

Mining Unexpected Temporal Associations: Applications in Detecting Adverse Drug Reactions*

Huidong Jin^{*,1} Jie Chen¹ Hongxing He¹ Graham Williams^{2,1}
Chris Kelman³ Christine M. O’Keefe¹

March 6, 2005

Abstract

In this paper we develop new data mining techniques for discovering valuable temporal episodes where some undesired outcome unexpectedly occurring after a certain event pattern, e.g., **esophagitis** occurring after taking **alendronate**. We define these episodes as Unexpected Temporal Association Rules (UTARs). Due to their *unexpectedness* and *infrequency*, existing techniques cannot effectively discover UTARs. To handle

*Prepared for submission to a data mining journal *IEEE Transactions on Knowledge and Data Engineering*.

¹CSIRO Mathematics and Information Sciences, GPO Box 664, Canberra ACT 2601, Australia.
Email:{Warren.Jin, Jie.Chen, Hongxing.He, Christine.OKeefe}@csiro.au. Corresponding author is Dr. Jin.
Tel: +61 2 62167258. Fax: +61 2 62167111.

²Australian Taxation Office, 51 Allara Street, Canberra ACT 2601, Australia. Email: Graham.Williams@togaware.com. The work was mainly done when Dr. Williams was with CSIRO.

³National Centre for Epidemiology and Population Health, the Australian National University, Canberra, ACT, 0200 Australia. Email: Chris.Kelman@anu.edu.au.

the unexpectedness we introduce a new interestingness measure, *residual-leverage*, for UTARs, and develop a novel case-based exclusion technique for its calculation. Combining it with an event-oriented data preparation technique to handle infrequency, we develop a new data mining algorithm, UTARM, to find the most interesting pairwise UTARs. We illustrate UTARM with a successful application on detecting Adverse Drug Reactions (ADRs) from large healthcare administrative databases. The algorithm efficiently and reliably short-lists ADRs such as **alendronate** unexpectedly being associated with **esophagitis**, **nefazodone** with **hepatitis**, **flucloxacillin** with **hepatitis**, and **ACE inhibitors** with **angioedema**. Compared with competitive algorithms, UTARM identifies potential ADRs much more effectively. Current results are very promising for post-market monitoring of prescription medicines.

Keywords: Data mining, temporal associations, unexpected temporal associations, sequential patterns, case-based exclusion, adverse drug reactions

1 Introduction

Throughout business, science, and engineering, a large number of events are recorded with corresponding temporal information (i.e., timestamps). We call such a sequence of events *a temporal event sequence*. Substantial research has focused on searching for interesting sequential patterns or temporal associations from a single or multiple event sequences [4, 16, 10, 13, 14, 22]. Most existing techniques focus on discovering frequent patterns/itemsets [11]. However, in many real-world applications we are more interested in finding *undesired episodes* — some outcome unexpectedly occurring after a certain event pattern, e.g., **esophagitis** occur-

ring after taking **alendronate**. Mining these episodes is of great value, particularly if outcomes are life threatening. Due to unexpectedness, such an episode may not necessarily indicate the event pattern confidently implies the outcome as indicated by *Temporal Association Rules* (TAR) [7, 10, 13, 14] or a *sequential patterns* [4, 16]. In addition, often such an episode does not occur so frequently, especially when the outcome is severe, due to regulatory processes such as pre and post market drug monitoring. Thus, it is impractical to identify these unexpected and infrequent episodes through traditional methods for finding sequential patterns and temporal associations — we require new techniques for finding these undesired and rare episodes. In this paper we introduce Unexpected Temporal Association Rules (UTARs) to describe these episodes, and provide a data mining algorithm for their discovery.

Adverse Drug Reactions (ADRs) provide a real-world example to illustrate where UTARs are of great value. The World Health Organisation (WHO) defines an ADR as *a response to a drug which is noxious and unintended, and which occurs at doses normally used or tested in man for the prophylaxis, diagnosis, or therapy of disease, or for the modification of physiological function* [2]. Informally, an ADR indicates an undesirable response associated with the use of a drug. It was estimated that in 1994 over 2 million patients hospitalised in USA had serious ADRs, which makes ADRs between the fourth and sixth leading cause of death in USA [12]. Besides the human suffering, this represents a tremendous impact on healthcare costs. The cost of drug-related problems, which ADRs constitutes a key part, was estimated to be about \$177.4 billion in the ambulatory setting in USA in 2000 [8]. However, studies also show that 30% to 60% of ADRs are preventable or avoidable by careful prescribing and monitoring [5]. Therefore, systematic post-market monitoring of ADRs is of financial and social importance.

Current drug monitoring systems have generally relied on spontaneous reporting systems like the Australian ADR Reporting System under which medical practitioners and others voluntarily report observed ADRs [1]. This method systematically underestimates the frequency of ADRs, typically by a factor of about 20 [6]. Unexpected drug side-effects might go unnoticed until a very substantial number of patients have been adversely affected. While manual *chart review* has been considered the “gold-standard” for identifying adverse events in the research setting, it is too costly for routine use. Thus, it is paramount to develop some techniques for automatic detection of ADRs at an earlier stage [17]. Current ADR detection systems search through various data sources for ADR “signals”, such as a drug-symptom pair (e.g., **alendronate** followed by **esophagitis**), to identify patients suffering ADRs [6, 17]. These systems can reduce costs significantly for specific types of ADRs providing the domain knowledge is given in advance. In contrast, our research aims to identify ADR signs (drug-disease/symptom pairs) as most interesting UTARs with little prior domain knowledge, by mining healthcare databases. Short-listed UTARs will help medical experts identify genuine ADRs more effectively and efficiently. Discovered ADR signs¹ can then be used in the current ADR detection systems to prevent or ameliorate further adverse events. Our work can help these systems move towards automation.

Another application of UTARs is in market basket problems containing temporal information. Marketing often uses sales promotions, such as discount coupons or special offers, to obtain short term increased sales. Such promotions can have an unexpected impact on other specific products. For example, brand A milk promotions may reduce sales of brand B milk.

¹For simplicity, we will use *an ADR*, rather than, *an ADR sign such as a drug-symptom pair*, to indicate *an adverse event potentially induced by specific drug usage* hereafter.

Marketers are very interested in identifying promotions (their own or their competitors') which unexpectedly lead to such decreases in sales of their own products. UTARs provide a mechanism to describe these undesired episodes too.

We characterise the concept of UTARs as follows. A number of subjects have their own sequence of events, each of which is a list of events together with their timestamps as exemplified in Fig. 1. We introduce a period length T to consider limited effect periods of related events. The aim is to find interesting patterns of events A and C such that A followed by C occurs relatively frequently within the T -sized periods but the occurrence of C is an unexpected outcome of the occurrence of A . Relative frequency is emphasised here because it is practicable to design algorithms to identify them from large real-world databases that commonly contain some noise, such as healthcare databases [27]. In addition, such kinds of patterns are applicable to more subjects, and are of more value. The discovered results, UTARs, can be expressed in the form of $A \xrightarrow{T} C$ in order to distinguish them from a traditional TAR denoted by $A \xrightarrow{T} C$. We read this as indicating that *the consequent* C occurs unexpectedly within a T -sized period after *the antecedent* A . We note the following points:

- The temporal constraint is important to ensure the sensibility of UTARs. The period length T constrains the temporal relation between A and C .
- The terms A and C may be either existence patterns (i.e., a set of events) or sequential patterns (i.e., an ordered list of events). For simplicity, we will only discuss existence patterns in this paper.
- The consequent C can be specified to restrict the search space for UTARs. The sequences

containing C are called *case sequences* while other sequences are called *noncase sequences*. Events that occur around C , constrained by T , are chosen for composing and evaluating the antecedent A . This is *event-oriented data preparation* and is intended to handle infrequency aspect of UTARs.

- The unexpectedness of UTARs is handled through the introduction of a new interestingness measure, the *residual-leverage*. We propose a *case-based exclusion technique* for its calculation for pairwise UTARs, which prunes the events that are *not* unexpectedly associated with C so as to highlight events unexpectedly associated with C . Based on these, we establish a new data mining algorithm, UTARM, to detect pairwise UTARs. However, our proposed techniques are extensible to longer patterns, such as diseases induced by drug-drug interactions.

We apply UTARM to detect ADRs from a large healthcare database. Known ADRs, such as alendronate \xrightarrow{T} esophagitis, nefazodone \xrightarrow{T} hepatitis, and ACE inhibitors \xrightarrow{T} angioedema, are successfully listed among the most interesting UTARs without prior knowledge. Of course, domain expertise is necessary on evaluating the results. Among these UTARs we also identify another ADR, flucloxacillin \xrightarrow{T} hepatitis. The UTARM’s performance is experimentally reliable with various parameter settings. Compared with OPUS_AR⁺ (extended from OPUS_AR [26]), we demonstrate that the proposed techniques can detect genuine ADRs more effectively. These findings are judged by domain experts to be of interest.

The remainder of the paper is organised as follows. Related work is discussed in Section 2. In Section 3, the problem of finding UTARs is formulated. In Section 4, we propose the algorithm UTARM to identify the most interesting UTARs based on residual-leverage. In Section 5, typical and comparative results are presented followed by the sensitivity analysis

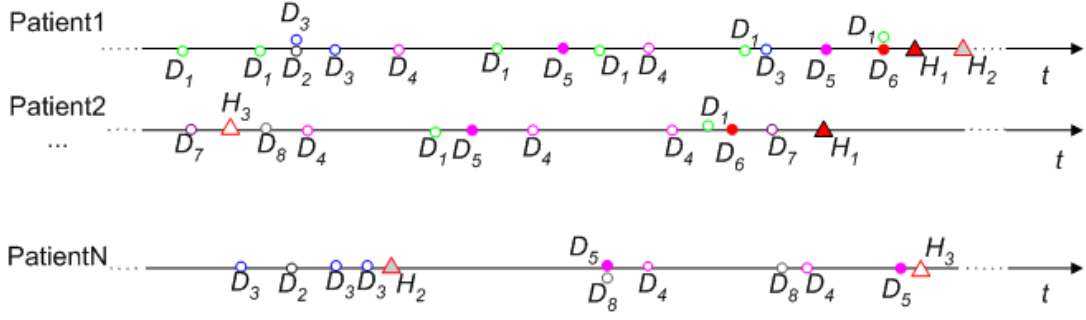


Figure 1: Illustration of temporal event sequences. D_i and H_j indicate different categories of events. E.g., D_i indicates taking drug i while H_j indicates an onset of disease j in healthcare data sets.

of the UTARM parameters. The research results are summarised and discussed in Section 6.

2 Related Work

Our work is closely related to the problem of mining TARs that has attracted much attention in recent years [20, 11]. Along this direction, Li et al. [14] studied TARs during time intervals specified by a user-given calendar schema. Lee et al. [13] explored the problem of mining TARs in publication databases where time intervals rather than timestamps were used. Harms and Deogun [10] also presented an efficient method for finding frequent TARs in one or more sequences that precede the occurrence of patterns in other sequences. Our proposed UTAR $A \xrightarrow{T} C$ indicates the consequent C unexpectedly occurs within a T -sized period after the antecedent A while both occur together at low-frequency. The UTAR is different from a traditional TAR where a consequent and an antecedent occur together frequently within a data subset specified by a temporal constraint.

Our pairwise UTARs can also be viewed as sequential patterns [4] (or episodes in [16])

where a collection of events occur relatively close together in a given partial order. Based on the Apriori property, Srikant and Rakesh developed a multiple-pass, candidate-generation-and-test approach, GSP, for searching sequential patterns whose support is larger than a user-specified threshold [21]. After that, many similar techniques have been developed. For example, Mannila *et al.* gave efficient algorithms for the discovery of all sequential patterns from a given class of patterns [16]. Han *et al.* developed a sequential mining approach, FreeSpan [9]. Its basic idea is to use frequent items to recursively project sequence databases into a set of smaller projected databases and grow subsequence fragments in each projected database. Our previous work [7] focused on finding patterns with high risks that occur frequently before a rare event from temporal event sequences. Sun *et al.* [22] studies patterns which happen frequently only in the periods prior to a target event over a long sequence. Recently, they discussed the discovery of negative event-oriented associations, in which the antecedent happens frequently all the time but before the consequent [23].

Clearly, all these techniques concentrate on finding frequent sequential patterns/itemsets. They are not suitable for identifying infrequent UTARs like ADRs, which is the aim of this research. The existing techniques consider different aspects of temporal data mining; we consider our research as complementary to them.

Wang *et al.* discussed how to embed the user knowledge in the association rule mining procedure, and established some promising techniques to find more unexpected associations of user interest [24]. In many areas, such as medication, there is too much domain knowledge to be considered. Furthermore, the user knowledge is not always easily available as pointed out in [24]. In contrast, our solution is to better use the data and automatically exclude the associations which are not interesting during the mining procedure.

3 Problem Formulation

Before introduction of UTARs to describe undesired episodes in temporal event sequences, we brief some techniques related with TARs.

As we know, an association rule is an implication in the form of $A \rightarrow C$, meaning that the presence of A implies the presence of C where A and C are mutually disjoint [28]. The effect of events usually last a short time period, e.g., drug usage will produce an adverse event within weeks to months, so temporal constraints need to be imposed. Many algorithms have been proposed to explore the problem of discovering temporal association relationships in database subsets, i.e., to determine association rules from a given subset of database specified by time constraints [13]. In this paper, we use the notation $A \xrightarrow{T} C$ to indicate explicitly that A and C occur within subsequences constrained by a certain time window T , i.e., C occurs within a T -sized period after A . Given the temporal constraint T , a TAR $A \xrightarrow{T} C$ holds in a set of sequences Θ if

1. its *support* $\text{supp}(A \xrightarrow{T} C)$, i.e., the proportion of T -constrained subsequences² in Θ which contain $A \cup C$ and in which A occurs prior to C at least once, is greater than a support threshold θ_s ;
2. its *confidence* $\text{conf}(A \xrightarrow{T} C) = \frac{\text{supp}(A \xrightarrow{T} C)}{\text{supp}(A \xrightarrow{T})}$, where $\text{supp}(A \xrightarrow{T})$ indicates the proportion of T -constrained subsequences² that contain A , is greater than a confidence threshold θ_c .

For example, for the three temporal event sequences shown in Fig. 1, if T is set as T_h as in Fig. 2, $\text{supp}(D_1 D_6 \xrightarrow{T} H_1) = \frac{2}{3}$ and $\text{conf}(D_1 D_6 \xrightarrow{T} H_1) = 1.0$.

As discussed in Section 1, undesired episodes occur normally at low frequency because of

²We choose at most one T -constrained subsequence for each temporal event sequence.

their unexpectedness and some regulatory processes. As a result, if we want to identify an undesired episode such as an ADR through finding infrequent TARs, the support threshold θ_s and the confidence threshold θ_c should be set very small. This will have innumerable valid TARs, and make the corresponding discovery algorithms unmanageable [26]. In addition, there remains the other problem of setting these threshold values properly. Thus, it is still too complicated to identify undesired episodes from TARs. The situation is similar for other temporal data mining models, e.g., sequential patterns [4, 9, 16] and event-driven sequential patterns [7, 22].

One method to identify undesired episodes from these innumerable TARs is to find the most interesting rules [19, 26]. For example, we may simply extend the OPUS_AR algorithm [26] to consider the temporal constraints and work on event-oriented data. We call it OPUS_AR⁺. It can return a prespecified number of TARs that maximise an association quality measure. However, the existing interestingness measures, such as support, confidence, risk ratio, odds ratio, and leverage, are not suitable for ranking TAR for detecting undesired episodes like ADRs.

As discussed above, due to their infrequency and unexpectedness, supports and confidences for TARs corresponding to undesired episodes are normally low. It seems impracticable to identify undesired episodes by only choosing TARs with high support (or confidence). The risk ratio and odds ratio are two measures of event effect which are appealing to domain experts, and are commonly used in the side-effect evaluation [7, 17]. For example, the risk ratio for $A \xrightarrow{T} C$ is $RR(A \xrightarrow{T} C) = \frac{supp(A \xrightarrow{T} C) / supp(A \xrightarrow{T})}{supp(\neg A \xrightarrow{T} C) / supp(\neg A \xrightarrow{T})}$, where $\neg A$ indicates that A does not occur. It describes to which degree the occurrence of A increases, with the T -constraint, the occurrence rate of C . However, both measures are not suitable for detecting undesired

episodes like ADRs from large databases. The first reason is that, due to noise or biases in data caused by data incompleteness and lack of reporting accuracy [27], the ranking based on these ratios may not be reliable. The second reason is that they have widely been used in the side-effect evaluation like pre-marketing drug screening, so that their values for ADRs in post-market healthcare databases are normally low. This, as also shown in Section 5, makes them inappropriate for highlighting genuine ADRs.

As a measure of the association strength, *leverage* indicates the proportion of sequences that exhibit the association in excess of those that would be expected when the antecedent A and the consequent C were independent of each other [26]. Considering the temporal constraint, we have

$$leverage(A \xrightarrow{T} C) = supp(A \xrightarrow{T} C) - supp(A \xrightarrow{T}) \times supp(\xrightarrow{T} C). \quad (1)$$

If we set $T = T_h$ for sequences having H_1 and $T = T_c$ for sequences without H_1 in Fig. 2, $leverage(D_1 \xrightarrow{T} H_1) = \frac{2}{3} - \frac{2}{3} * \frac{2}{3} = \frac{2}{9}$.

Using Eq.(1) and noting that $0 \leq supp(\xrightarrow{T} C), supp(A \xrightarrow{T}) \leq 1$, we may have the following inequality between the confidence and leverage.

$$\begin{aligned} conf(A \xrightarrow{T} C) &= \frac{supp(A \xrightarrow{T} C)}{supp(A \xrightarrow{T})} = \frac{leverage(A \xrightarrow{T} C)}{supp(A \xrightarrow{T})} + supp(\xrightarrow{T} C) \\ &\geq leverage(A \xrightarrow{T} C). \end{aligned} \quad (2)$$

Unfortunately, leverage can not express the unexpectedness that we need for detecting undesired episodes. For example, based on the leverage values, **paracetamol** (N02BE01 in Tables 2 and 6 in Section 5.2) is strongly associated with two diseases **esophagitis** and **angioedema**. Paracetamol is a medicine that is used for mild to moderate pain and fever, and it may be viewed as a therapeutic drug for both diseases, rather than inducing the diseases

as adverse reactions.

Thus, we need a new measure in order to distinguish, from TARs, Unexpected TARs (UTARs) like ADRs. For clear indication of the unexpectedness, we denote an UTAR as $A \xrightarrow{T} C$. We define a new interestingness measure, residual-leverage, for UTARs as follows.

Definition 1 The *residual-leverage*³ of the UTAR $A \xrightarrow{T} C$ is

$$resilev(A \xrightarrow{T} C) = supp(A \xrightarrow{T} C) - supp(A \xrightarrow{T}) \times supp(\xrightarrow{T} C) \quad (3)$$

where $supp(A \xrightarrow{T})$ is the proportion of T -constrained subsequences that unexpectedly contain A and $supp(A \xrightarrow{T} C)$ is the proportion of T -constrained subsequences that unexpectedly contain A followed by C .

We explicitly embed the unexpectedness of an UTAR into its supports. Different from the traditional support, only the subsequences which contain A unexpectedly will contribute the support in the definition of residual-leverage. For example, for Case 1 in Fig. 2, D_1 occurs within T_h -period prior to and also far before the unique H_1 , and it is difficult to say that this sequence support the UTAR, $D_1 \xrightarrow{T_h} H_1$. Clearly, we have

$$supp(A \xrightarrow{T}) \leq supp(A \xrightarrow{T}), \text{ and} \quad (4)$$

$$resilev(A \xrightarrow{T} C) \leq supp(A \xrightarrow{T} C) \leq supp(A \xrightarrow{T} C). \quad (5)$$

³Ideally, it should be defined as $supp(A \xrightarrow{T} C) - supp(A \xrightarrow{T}) \times supp(\xrightarrow{T} C)$. Here $supp(\xrightarrow{T} C)$ is the proportion of T -constrained subsequences that unexpectedly contain C in all subsequences. Note that the consequent C occurs very infrequently, and especially with little possibility occurs more than once in one event sequence. This makes it arduous, if not impossible, to judge whether an event sequence *unexpectedly* contain C simply based on data. Dislike the calculation of $supp(A \xrightarrow{T})$, the calculation of $supp(\xrightarrow{T} C)$ is not easy. Thus, we roughly replace $supp(\xrightarrow{T} C)$ with $supp(\xrightarrow{T} C)$ in Eq.(3).

Based on our new case-based exclusion technique, we will give a simple method to calculate $supp(A \xrightarrow{T})$ and $supp(A \xrightarrow{T} C)$ in Section 4.

Besides the support and confidence thresholds, we can set the third threshold, say, θ_r , for residual-leverage to define a valid UTAR. This threshold greatly increases the likelihood of finding interesting UTARs. Again, the number of valid UTARs found can be too large if the threshold θ_r is set too low. Or, some really interesting UTARs become invalid if the threshold is set too high. To avoid the problem of setting thresholds properly, we would like to select a prespecified number of, say 10, UTARs that have as high residual-leverage as possible. Another reason to rank UTARs only according to residual-leverage is that a large residual-leverage also indicates a large support and confidence in the sense that $supp(A \xrightarrow{T} C) \geq resilev(A \xrightarrow{T} C)$ and $conf(A \xrightarrow{T} C) \geq resilev(A \xrightarrow{T} C)$. These inequalities are indicated by Eqs. (2)-(5) above and Theorem 1 in Section 4.

4 Searching for Unexpected Temporal Association Rules

In this section, we develop a simple but effective algorithm to search for interesting UTARs. We concentrate on pairwise UTARs such as an ADR where one single drug might induce one disease, i.e., **drug A** \xrightarrow{T} **disease C**. It is mainly because pairwise UTARs are of great value, and some successful experience on them can pave the way for us to discover more sophisticated UTARs in the future. Our algorithm, UTARM, is outlined in Algorithm 1. We take the sequences illustrated in Fig. 2 as examples to explain UTARM.

First, we initialise some parameters that are explained as follows.

- The consequent C is specified to restrict the search space. In Fig. 2, e.g., H_1 is specified.

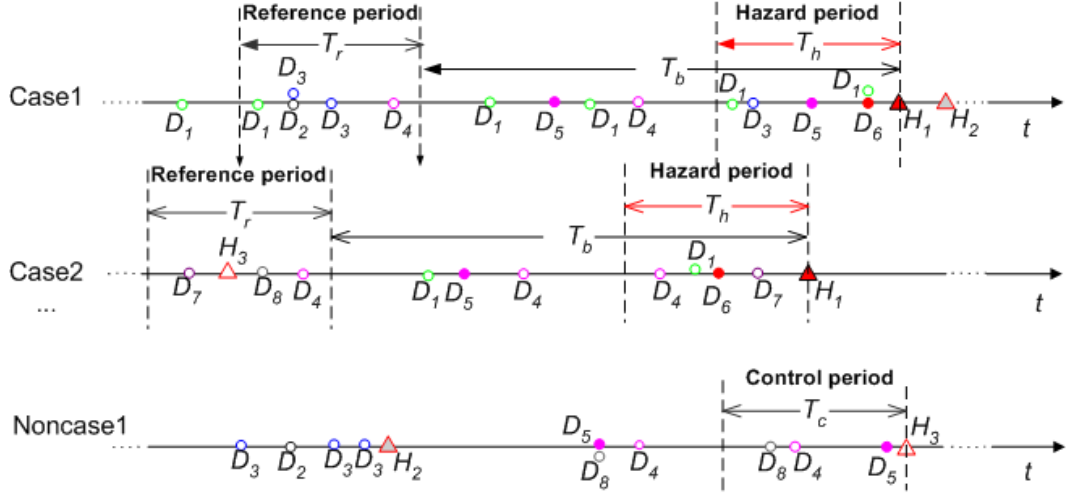


Figure 2: Illustration of the event-oriented data preparation and case-based exclusion. D_i and H_j indicate a drug-taking event and a disease occurrence respectively. H_1 indicates the consequent event in this example.

- A study period is determined by $[t_S, t_E]$. A case sequence whose first occurrence of the consequent C is not in the study period will not be considered. We concentrate on the first occurrence so as to facilitate the implementation of the case-based exclusion and choose at most one T -constrained subsequence for each case sequence.
- The time lengths T_h , T_r , T_b , and T_c indicate, as illustrated in Fig. 2, *the hazard period length, the reference period length, the period length between the hazard and the reference periods, and the control period length* respectively.

We set the hazard period for each case sequence as the T_h -sized period before the first occurrence of the consequent C which occurs in the study period. Clearly, the events during the hazard period may associate with the consequent C . Among them, some are less possible to unexpectedly lead to C . If domain experts can provide a list of events which do not unexpectedly lead to C , we may simply exclude them from the hazard period. However, this

domain knowledge is often unavailable or out of date [24]. However, for each case sequence, we can use the events which occurred before the hazard period to deduce some events which do not unexpectedly lead to C . For example, if an event occurs repeatedly in the case sequence, say, if the patient often takes one drug like D_1 for Case 1 in Fig. 2, this event is less likely to induce the first occurrence of the consequent C . The underlying reason is that most events, such as taking a drug, have short-term effects [25], and the subject has similar response to a certain event. Thus, we may disregard this event in the hazard period, and deduce that the remaining ones are more likely to induce C unexpectedly. This exclusion operation is carried out only based on a single case sequence, and is termed as the *case-based exclusion*. To further simplify this exclusion operation, we borrow the concept of the reference period from the case-crossover study [15, 25]. The reference period is a T_r -sized period which is T_b -sized interval before the hazard period as illustrated in Fig. 2. If the event, e.g., taking a drug, in the reference period is protective/therapeutic to C , the patient is not surprised to see the occurrence of the consequent C . In contrast, the patient (or his/her doctors) will believe the events within the reference period are basically “safe” to her/him because there is no occurrence of an unexpected outcome like C , even long after the reference period. Thus, the events within the reference period are not unexpected to the case with respect to the consequent C , and they can be excluded for mining UTARs. For example, for Case 1 in Fig. 2, D_1 - D_4 are in the reference period, and D_1 , D_3 , D_5 , and D_6 in the hazard period. D_1 and D_3 are excluded, and only D_5 and D_6 are kept for the subsequence for Case 1. Similarly, D_1 and D_6 are left for Case 2.

In Step 3 of UTARM, for each noncase sequence, we have two different methods to set a T_c -sized control period to choose the noncase subsequence. We may randomly choose the

Algorithm 1 *UTAR Mining (UTARM)*

1. *Initialise parameters, such as, the consequent C , the study period $[t_S, t_E]$, time period lengths T_h , T_r , T_b , and T_c , and the number of output UTARs k ;*
2. *Prepare case subsequences from case sequences which have the first occurrence of C during the study period: choose events from the hazard period, and exclude some of them based on the case-based exclusion with respect to the consequent C ;*
3. *Choose noncase subsequences from the control period from noncase sequences;*
4. *Calculate supports and residual-leverage of each event;*
5. *Rank the events in the descending order of residual-leverage, and return the top k interesting UTARs.*

control period within $[t_S - T_c, t_E]$. The other one is motivated by the matched case-control study [18] in order to avoid possible impact from other factors such as age and gender. We may set the control period to match a case according to, say, demographic data. A noncase is chosen from the same demographic, e.g., age-gender, group as the case. In addition, the noncase has an event similar with C which also occurred closely to C in the case, say, the onset of another disease in the same month of the onset of the disease C . For example, Noncase 1 in Fig. 2 has H_3 which occurs in the same month with H_1 of Case 1, thus the T_c -sized period before H_3 is set as the matched control period for Noncase 1. The events within the control period, say, D_4 , D_5 , and D_8 for Noncase 1, compose the noncase subsequence.

In Step 4, putting these case and noncase subsequences together, we then calculate supports for events that appeared in the hazard periods which are based on the case subsequences after exclusion and the noncase subsequences. Then, we compute the residual-leverage value of each event according to Eq.(3). For example, for the three sequences in Fig. 2, $resilev(D_1 \xrightarrow{T} H_1) = supp(D_1 \xrightarrow{T} H_1) - supp(D_1 \xrightarrow{T}) \times supp(\xrightarrow{T} H_1) = \frac{1}{3} - \frac{1}{3} * \frac{2}{3} = \frac{1}{9}$

and $resilev(D_6 \xrightarrow{T} H_1) = \frac{2}{3} - \frac{2}{3} * \frac{2}{3} = \frac{2}{9}$. Thus, residual-leverage roughly indicates the situation where D_6 is more unexpectedly associated with H_1 than D_1 .

Finally, the algorithm outputs the k most interesting events. Together with the consequent C , we have the k most interesting UTARs.

In UTARM, the hazard and the control periods are set according to the consequent C , and restrict the events in the calculation of supports. This *event-oriented data preparation* makes it possible to detect the usually infrequent UTARs. The *case-based exclusion* in Step 2 is simple and easy-implemented, but it plays a key role in UTARM. Without this operation, UTARM will perform the same as OPUS-AR⁺. The following theorem indicates that, using the case-based exclusion, the residual-leverage value is not greater than the leverage value.

Theorem 1 *With the case-based exclusion, $resilev(A \xrightarrow{T} C) \leq leverage(A \xrightarrow{T} C)$.*

Proof: Let the proportion of case subsequences where A is excluded be $supp_{ex} (\geq 0)$. Because of the case-based exclusion, we get

$$supp(A \xrightarrow{T} C) - supp(A \xrightarrow{\leftarrow T} C) = supp_{ex}, \quad (6)$$

$$supp(A \xrightarrow{T}) - supp(A \xrightarrow{\leftarrow T}) = supp_{ex}. \quad (7)$$

Considering the definition of support, we get another inequality $0 \leq supp(\xrightarrow{T} C) \leq 1.0$.

Simply, according to the definitions of residual-leverage and leverage, we get

$$\begin{aligned} leverage(A \xrightarrow{T} C) - resilev(A \xrightarrow{\leftarrow T} C) &= supp(A \xrightarrow{T} C) - supp(A \xrightarrow{\leftarrow T} C) \\ &= \left(supp(A \xrightarrow{T}) - supp(A \xrightarrow{\leftarrow T}) \right) \times supp(\xrightarrow{T} C) \\ &= supp_{ex} - supp_{ex} * supp(\xrightarrow{T} C) \geq 0. \end{aligned} \quad (8)$$

The proof is completed. ■

Combining Theorem 1 with Eq.(2), we have $\text{conf}(A \xrightarrow{T} C) \geq \text{resilev}(A \xrightarrow{T} C)$. That means a large residual-leverage value indicates a large confidence value. For example, if $\text{resilev}(A \xrightarrow{T} C) > \theta_r$, then $\text{leverage}(A \xrightarrow{T} C) > \theta_r$. Thus, as discussed in Section 3, we may simply choose the most interesting UTARs only based on residual-leverage, and avoid the problem of setting appropriate thresholds for confidence as well as support.

5 Experimental Setting and Results

5.1 The QLDS

The Queensland Linked Data Set (QLDS) has been used as a test data set for our algorithm. This data set has been made available to CSIRO under an agreement between the Queensland Department of Health and the Commonwealth Department of Health and Ageing (DoHA). This data set links de-identified patient level hospital morbidity data (for the period between 1 July 1995 and 30 June 1999), Medicare Benefits Scheme (MBS) data, and Pharmaceutical Benefits Scheme (PBS) data (for the period from 1 January 1995 to 31 December 1999) in Queensland, Australia [27]. For convenience, we refer to 1 January 1995 as the first day hereinafter. Thus the time period for the whole data set is [1, 1826].

Each record in the hospital data corresponds to one inpatient episode, and each disease is coded in the ICD9-CM system, such as 530 for **esophagitis**. Each record in the MBS data corresponds to a medical service for one patient. Similarly, each record in the PBS data corresponds to one prescription supplied to one patient, and the 3842 distinct description items are mapped into 758 distinct codes in the WHO ATC code system. Each ATC code is a structured 7 digit alpha-numeric code with 5 levels, e.g., the ATC codes for **alendronate** ⁴,

⁴Alendronate is an aminobisphosphonate which specifically inhibits osteoclast-mediated bone resorption. It was approved for treatment of osteoporosis in postmenopausal women and Paget’s disease of bone [2].

nefazodone ⁵, and ACE inhibitors ⁶ are M05BA04, N06AX06, and C09AA?? ⁷ respectively [3].

As a result, a patient may have more than one hospitalisation event, or MBS or PBS record. The QLDS contains records for 1,176,294 individuals who were hospitalised in Queensland during 1 July 1995 to 30 June 1999 [27]. For each individual, a temporal event sequence is generated to record sequentially his/her hospitalisation events, PBS and MBS events, as well as their timestamps.

5.2 Typical Experimental Results

We first describe some typical results generated by UTARM, in comparison with OPUS_AR⁺, a simple extension of OPUS_AR [26] for temporal data. We concentrate on three types of diseases, **esophagitis**, **hepatitis**, and **angioedema**, which are suspected to be sometimes induced by **alendronate**, **nefazodone**, and **ACE inhibitors** respectively [2]. Like other data mining results, it is unrealistic to expect every interesting UTAR found to be appealing to domain experts. This is partially because the databases often contain noise [27]. A realistic goal is to, via better using databases, reliably put these ADRs in the 10 or 20 most interesting UTARs. Domain experts still need to examine the (but much fewer) short-listed UTARs to identify the genuine ADRs.

As for the parameters of UTARM, we set $T_h = 180$, $T_r = 3 \times T_h$, $T_c = T_h$, and $T_b = 365$ in days by default. T_b is set as one year in order to eliminate some seasonality of drug usage. We set the study period as [730, 1645] where 1645 is biggest for all possible hospitalisation

⁵Nefazodone (Serzone) is one of the new antidepressants. It is related to the SSRIs (Selective Serotonin Re-uptake Inhibitors, a widely used group of antidepressants), but has a dual action in that it works on both sides of the serotonin synapse [2].

⁶ACE inhibitors (Angiotensin Converting Enzyme inhibitors) are a commonly used treatment of high blood pressure and heart disease [2].

⁷Here ‘?’ is a wildcard. There are about 9 different ATC codes, for ACE inhibitors, from C09AA01 to C09AA10 except C09AA07. For example, C09AA03 represents Lisinopril [3]

Table 1: The top 10 drugs unexpectedly associated with **esophagitis** short-listed by UTARM for older-female patients (The *italic* row is for the suspected drug **alendronate**).

Rank in		Drug in ATC code	Drug name	Resi -lev	Leve -rage	$supp(A \xrightarrow{T})$ $\times N (=58172)$	$supp(A \xrightarrow{T} C)$ $\times N (=58172)$	Risk ratio
Resilev	Leverage							
1	4	A03FA02	Cisapride	4.37E-3	1.25E-2	1790	534	3.245
2	15	J07AL01	Pneumococcal	3.79E-3	3.92E-3	1854	510	1.829
3	13	A02BA04	Nizatidine	2.48E-3	4.06E-3	995	300	2.434
4	31	<i>M05BA04</i>	<i>Alendronate</i>	<i>1.51E-3</i>	<i>2.28E-3</i>	<i>628</i>	<i>186</i>	<i>2.279</i>
5	2	A02BC01	Omeprazole	1.50E-3	1.55E-2	1528	326	3.692
6	44	C09CA04	Irbesartan	1.50E-3	1.54E-3	658	190	1.886
7	35	B01AC06	Aspirin	1.43E-3	1.85E-3	1035	245	1.667
8	14	A03FA01	Metoclopramide	1.31E-3	4.00E-3	2288	434	1.645
9	49	C10AA05	Atorvastatin	1.25E-3	1.47E-3	763	192	1.718
10	42	A03FA03	Domperidone	7.75E-4	1.56E-3	524	127	2.023

events in the QLDS. We partitioned all the patients into 6 age-gender groups according to their demographic data. Ages were discretised into young ($[0, 20)$), middle-aged ($[20, 59]$), and older (≥ 60) groups. For **esophagitis**, we focused on the older-female group because 9092 out of 33094 **esophagitis** patients are female and older. For each older-female **esophagitis** patient, we chose up to 6 (only 5 are available for some of them) different matched non-**esophagitis** patients who were from the same age-gender group as the **esophagitis** patient and hospitalised due to some disease in the same month of the **esophagitis** onset. There were 49080 non-**esophagitis** subsequences. Then the total number of subsequences $N = 58172$. For 2912 **hepatitis** patients, we were interested in 1034 middle-aged-female **hepatitis** patients. We chose about 21 matched non-**hepatitis** subsequences for each of them and got 21660 non-**hepatitis** subsequences in total. Then the total number of subsequences $N = 22694$. For 286 **angioedema** patients, we were interested in 75 **angioedema** patients older than 60, and chose about 21 matched non-**angioedema** patients for each of them. There were 1515 non-**angioedema** patients.

Table 1 lists the 10 drugs which have the highest residual-leverage values with respect

Table 2: Some drugs associated with **esophagitis** for older-female patients.

Rank in		Drug in ATC code	Drug name	Leve -rage	Resi -lev	Risk ratio	$supp(A \xrightarrow{T})$	$supp(A \xrightarrow{T} C)$
Leverage	Resilev						$\times N (=58172)$	$\times N (=58172)$
1	488	A02BA02	Ranitidine Hydrochloride	2.03E-2	-2.32E-3	2.263	8223	2468
2	5	A02BC01	Omeprazole	1.55E-2	1.50E-3	3.692	2491	1289
3	545	J07BB02	Influenza Vaccine	1.42E-2	-1.63E-2	1.708	10128	2407
4	1	A03FA02	Cisapride	1.25E-2	4.37E-3	3.245	2348	1092
5	402	A02BA03	Famotidine	1.03E-2	-4.76E-4	2.142	3848	1198
6	546	N02BE01	Paracetamol	8.46E-3	-2.93E-2	1.282	17095	3164
7	433	A02AD	Aluminium Hydroxide	6.71E-3	-7.46E-4	1.917	3006	860
8	530	C01DA02	Glyceryl Trinitrate	5.49E-3	-5.98E-3	1.406	5808	1227
9	535	R03AC02	Salbutamol	4.33E-3	-7.37E-3	1.334	5483	1109
10	537	N02AA59	Codeine with Paracetamol	4.26E-3	-7.53E-3	1.274	6732	1300
...
31	4	M05BA04	alendronate	2.28E-3	1.51E-3	2.279	681	239

to **esophagitis** generated by UTARM for the older-female patients. After the case-based exclusion, there are only 186 ($= supp(A \xrightarrow{T} C) \times N$) **esophagitis** patients from 628 ($= supp(A \xrightarrow{T}) \times N$) **alendronate** drug users. Comparing the supports listed in Table 2 where some drugs identified by OPUS_AR⁺ are ranked based on leverage, we can observe there is only 53 ($= 239 - 186$) patients taking **alendronate** within their reference periods. The UTAR **alendronate** \xrightarrow{T} **esophagitis** has the residual-leverage of 1.51×10^{-3} , and based on this, is ranked as 4 among 758 different kinds of drugs. As a comparison, the association rule **alendronate** \xrightarrow{T} **esophagitis** is ranked as low as 31 in Table 2. This is partially because there are only 239 older-female patients who took **alendronate** within their hazard periods, which is much smaller than other drug users in the table. The support for this association rule is 0.32%, and it is ranked as low as 91. Though its confidence is as high as 35.1% (due to the event-oriented data preparation), this association rule is ranked as low as 43 based on the confidence value. Its risk ratio is 2.279. It means an older female, if taking **alendronate**, is 2.279 times more likely to suffer from **esophagitis**. However, the rank based

Table 3: The top 10 drugs unexpectedly associated with hepatitis short-listed by UTARM for the middle-aged-female patients (The *italic* and the *slanted* rows are for nefazodone and flucloxacillin, respectively.).

Rank in		Drug in ATC code	Drug name	Resi -lev	Leve -rage	$supp(A \xrightarrow{T})$	$supp(A \xrightarrow{T} C)$	Risk ratio
Resilev	Leverage					$\times N(=22694)$	$\times N(=22694)$	
1	7	N05CD02	Nitrazepam	5.03E-4	1.22E-3	276	24	3.156
2	15	N02AC	Diphenylpropylamine Derivatives	4.30E-4	6.40E-4	49	12	7.008
3	25	<i>N06AX06</i>	<i>Nefazodone</i>	<i>3.78E-4</i>	<i>3.78E-4</i>	<i>53</i>	<i>11</i>	<i>4.593</i>
4	20	<i>J01CF05</i>	<i>Flucloxacillin</i>	<i>3.51E-4</i>	<i>5.19E-4</i>	<i>352</i>	<i>24</i>	<i>1.746</i>
5	19	P01BC01	Quinine Bisulphate	3.30E-4	5.40E-4	187	16	2.429
6	3	N05BA04	Oxazepam	3.26E-4	3.23E-3	562	33	3.826
7	10	N06AB05	Paroxetine	3.11E-4	8.99E-4	394	25	2.140
8	21	N06AG02	Moclobemide	3.05E-4	5.15E-4	375	24	1.694
9	26	C02AC01	Clonidine	2.94E-4	3.78E-4	51	9	4.593
10	12	N06AB06	Sertraline	2.66E-4	6.87E-4	504	29	1.691

on the risk ratio is 44. Thus, the residual-leverage can highlight ADRs much better than other measures. Similar situation can be observed for the other two ADRs. On the other hand, paracetamol (N02BE01) is ranked as 6 based on its leverage value, and is thought to be strongly associated with esophagitis as shown in Table 2. According to residual-leverage, it is ranked as low as 546. Note that similar interesting situation happens on several therapeutic drugs like ranitidine hydrochloride, famotidine and aluminium hydroxide. Furthermore, UTARM ran 57.2 seconds to generate Table 2 which is only 3.3 seconds longer than OPUS-AR⁺. These comparison results empirically support that the proposed measure, residual-leverage, can help us to remove some protective/therapeutic drugs⁸ automatically and efficiently for ADRs detection.

⁸Besides the data quality issue and the simple design for calculation of supports in UTARM, there exists another interesting reason for UTARM not able to remove all protective/therapeutic drugs: “treatment failures” cannot be distinguished from ADRs only based on data without priori knowledge. For example, the pro-motility drugs, including cisapride, metoclopramide, and domperidone in Table 1, are reserved either for patients who do not respond to other treatments or are added to enhance other treatments for gastroesophageal reflux disease (see http://www.medicinenet.com/gastroesophageal_reflux_disease_gerd/page6.htm). Thus, it is not surprised to see there are still many patients suffering esophagitis after taking these drugs. These are not ADRs but “treatment failures”.

Table 4: Some drugs associated with hepatitis for the middle-aged-female patients (The *italic* and the *slanted* rows are for nefazodone and flucloxacillin, respectively.).

Rank in		Drug in ATC code	Drug name	Leve -rage	Resi -lev	Risk ratio	$supp(A \xrightarrow{T})$	$supp(A \xrightarrow{T} C)$
Leverage	Resilev						$\times N(=22694)$	$\times N(=22694)$
1	36	N05BA01	Diazepam	5.39E-3	4.64E-5	3.910	1109	173
2	17	N05CD07	Temazepam	3.74E-3	4.75E-5	2.671	1296	144
3	6	N05BA04	Oxazepam	3.23E-3	5.80E-5	3.826	631	102
4	305	N02AA59	Codeine with paracetamol	3.06E-3	2.13E-5	1.727	2557	186
5	20	A03FA01	Metoclopramide hydrochloride	1.46E-3	4.83E-5	1.853	919	75
6	276	J01DA01	Cephalexin	1.36E-3	3.87E-5	1.390	1958	120
7	1	N05CD02	Nitrazepam	1.22E-3	8.89E-5	3.156	293	41
8	306	J01CA04	Amoxycillin	1.13E-3	2.08E-5	1.247	2643	146
9	303	R03AC02	Salbutamol	9.84E-4	2.12E-5	1.300	1814	105
10	7	N06AB05	Paroxetine	8.99E-4	6.30E-5	2.140	408	39
...
20	4	<i>J01CF05</i>	<i>Flucloxacillin</i>	<i>5.19E-4</i>	<i>6.81E-5</i>	<i>1.746</i>	<i>356</i>	<i>28</i>
...
25	3	<i>N06AX06</i>	<i>Nefazodone</i>	<i>3.78E-4</i>	<i>2.44E-4</i>	<i>4.593</i>	<i>53</i>	<i>11</i>

Table 3 lists the 10 drugs having the highest residual-leverage values with respect to hepatitis generated by UTARM, while Table 4 gives some drugs strongly associated with hepatitis generated by OPUS_AR⁺. 11 patients suffer hepatitis soon after taking nefazodone. Thus, the support for the UTAR nefazodone \xrightarrow{T} hepatitis is as low as 0.05%. Its residual-leverage is 3.78×10^{-4} , and it is ranked as 3 among 758 different kinds of drugs. As a comparison, OPUS_AR⁺ ranks the association rule nefazodone \xrightarrow{T} hepatitis as low as 25 based on leverage. It is worth pointing out that Row 4 in Table 3 indicates another very interesting ADR flucloxacillin \xrightarrow{T} hepatitis. We had not been aware of this ADR before we asked medical experts to examine the experimental results. According to the Australian Adverse Drug Reactions Bulletin [2], flucloxacillin is the most commonly reported to the Adverse Drug Reactions Advisory Committee (ADRAC) in association with hepatic reaction up to March, 1996⁹. As a comparison, OPUS_AR⁺ ranks this association rule as 20 in Table 4. On

⁹See <http://www.tga.gov.au/docs/html/aadrbltn/aadr9605.htm#liver>

Table 5: The top 10 drugs unexpectedly associated with **angioedema** short-listed by UTARM for the older-female patients (The three *italic* rows are for three **ACE inhibitors**.).

Rank in		Drug in ATC code	Drug name	Resi -lev	Leve -rage	$supp(A \stackrel{T}{\rightarrow})$ $\times N(=1590)$	$supp(A \stackrel{T}{\rightarrow} C)$ $\times N(=1590)$	Risk ratio
Resilev	Leverage							
<i>1</i>	<i>2</i>	<i>C09AA03</i>	<i>Lisinopril</i>	<i>4.49E-3</i>	<i>5.68E-3</i>	<i>82</i>	<i>11</i>	<i>3.759</i>
2	12	G03CA01	Ethinylloestradiol	3.45E-3	3.45E-3	11	6	12.48
3	22	R06AD02	Promethazine	2.98E-3	2.98E-3	48	7	3.306
4	20	J01EA01	Trimethoprim	2.50E-3	3.10E-3	43	6	3.616
5	33	M02AC	Methyl Salicylate	2.37E-3	2.37E-3	26	5	4.296
6	17	J01DA08	Cefaclor	2.12E-3	3.32E-3	77	7	2.608
7	44	J01CA04	Amoxycillin	2.05E-3	2.05E-3	143	10	1.556
<i>8</i>	<i>29</i>	<i>C09AA04</i>	<i>Perindopril</i>	<i>1.99E-3</i>	<i>2.59E-3</i>	<i>39</i>	<i>5</i>	<i>3.369</i>
<i>9</i>	<i>48</i>	<i>C09AA09</i>	<i>Fosinopril</i>	<i>1.90E-3</i>	<i>1.90E-3</i>	<i>42</i>	<i>5</i>	<i>2.632</i>
10	49	A01AB04	Amphotericin	1.89E-3	1.89E-3	21	4	4.209

the other hand, UTARM and OPUS-AR⁺ spent 18.6 and 18.3 seconds respectively. These comparisons again substantiates that UTARM can efficiently help us focus only on these most suspected drugs.

Table 5 lists the top 10 suspected drugs which unexpectedly leads to **angioedema** generated by UTARM, while Table 6 lists the top 10 drugs strongly associated with **angioedema** generated by OPUS-AR⁺. Two algorithms spent about 1.9 and 1.8 seconds, respectively. Interestingly, there is only one **ACE inhibitor** in Table 6, **Lisinopril** (C09AA03), among the top 10 drugs strongly associated with **angioedema**. However, three from 9 distinct **ACE inhibitors**, i.e., **Lisinopril** (C09AA03), **Perindopril** (C09AA04), and **Fosinopril** (C09AA09) are within the top 10 drugs short-listed by UTARM. They are ranked as 1, 8, and 9 respectively. In addition, these ranks advance much in comparison with those in terms of leverage. For example, the rank of **Fosinopril** (C09AA09) improves from 48 to 9. Even for **Lisinopril** (C09AA03), its rank advances from 2 to 1. Again **paracetamol** (N02BE01) is ranked as 9 based on leverage, while 211 based on residual-leverage. These comparisons again substantiates that the interestingness measure, residual-leverage, can help us to remove some therapeutic/protective

Table 6: Top drugs associated with **angioedema** for the older-female patients (The *italic* row is for a suspected drug **Lisinopril**).

Rank in		Drug in ATC code	Drug name	Leve -rage	Resi -lev	Risk ratio	$supp(A \xrightarrow{T})$	$supp(A \xrightarrow{T} C)$
Leverage	Resilev						$\times N(=1590)$	$\times N(=1590)$
1	42	C03CA01	Frusemide	9.16E-03	7.71E-04	3.270	200	24
2	1	<i>C09AA03</i>	<i>Lisinopril</i>	<i>5.68E-03</i>	<i>4.49E-03</i>	<i>3.759</i>	<i>84</i>	<i>13</i>
3	199	C01DA02	Glyceryl Trinitrate	5.41E-03	-1.78E-03	2.475	157	16
4	210	J07BB02	Influenza vaccine	5.29E-03	-4.30E-03	1.731	373	26
5	193	G03CA57	Oestrogens conjugated	5.20E-03	-1.38E-03	3.643	79	12
6	93	N05CD02	Nitrazepam	4.48E-03	2.84E-04	3.856	61	10
7	204	R03AC02	Salbutamol	4.02E-03	-2.57E-03	2.171	140	13
8	100	A02AD	Aluminium hydroxide	3.86E-03	2.61E-04	2.829	82	10
9	211	N02BE01	Paracetamol	3.78E-03	-9.37E-03	1.436	466	28
10	18	J01FA06	Roxithromycin	3.65E-03	1.26E-03	2.312	110	11

drugs automatically and let medical experts focus only on these most suspected drugs.

5.3 Sensitivity Examination

Now we show the proposed algorithm can reliably list ADRs in the most interesting UTARs with different parameter settings. For simplicity, we take the two known ADRs, **alendronate** \xrightarrow{T} **esophagitis** and **nefazodone** \xrightarrow{T} **hepatitis**, as examples to present experimental results.

Besides using the certain age-gender groups as in Section 5.2, we also examined our algorithm on all the groups for a given disease as shown in the first column of Table 7. We used two different ways to choose the noncase subsequences: 1) **Using matched control period**: for each case, about a prespecified number, say 6 for **esophagitis** patients and 21 for **hepatitis** patients, of different noncases are selected who suffer a disease onset in the same month of the first occurrence of the consequent in the case and are in the same age-gender group as the case; 2) **Using random control period**: where a T_c -sized subsequence is chosen randomly from each noncase sequence within $[t_S - T_c, t_E]$. Various hazard period

Table 7: Ranks of the suspected drugs for 18 different parameter settings.

Cases	Control period setting	Hazard period (in days)	Study period (in days)	Rank			
				$\text{alendronate} \xrightarrow{T} \text{esophagitis}$		$\text{nefazodone} \xrightarrow{T} \text{hepatitis}$	
				Resilev	Leverage	Resilev	Leverage
age-gender group: older-female for esophagitis, middle-aged -female for hepatitis	matched control period	180	910-1645	7	30	1	16
			730-1645	4	31	3	25
			550-1645	8	35	6	38
		120	730-1645	7	32	7	30
		90	730-1645	5	30	6	33
		60	730-1645	5	27	6	33
	random control period	180	910-1645	8	41	3	22
			730-1645	7	37	7	36
			550-1645	9	30	8	45
All groups	matched control period	180	910-1645	18	74	10	38
			730-1645	16	69	14	48
			550-1645	19	76	17	60
		120	730-1645	17	68	17	45
		90	730-1645	17	69	18	51
		60	730-1645	16	62	19	53
	random control period	180	910-1645	18	96	12	36
			730-1645	20	106	18	47
			550-1645	19	119	20	62

lengths T_h were tested such as 180, 120, 90, and 60 days. We also set different study periods such as from 550 to 1645, 730 to 1645, and 910 to 1645 as shown in the 4th column of Table 7. Note $T_b = 365$ by default, 550 was the minimal value we might choose in order to have at least a half year period available for the reference period.

For the older-female patients, the ranks of **alendronate** based on residual-leverage with respect to **esophagitis** range from 4 to 9 for the 9 different parameter settings. The average rank is 6.7. Its ranks based on leverage range from 27 to 41, which lie at least 21 behind. On average, they lie 25.9 behind. Similarly, for the middle-aged-female patients, the ranks of **nefazodone** \xrightarrow{T} **hepatitis** generated by UTARM range from 1 to 8. The average is 5.2. The ranks generated by OPUS_AR⁺ are from 16 to 45, and UTARM sorts the ADR 15 to 37 higher than OPUS_AR⁺ does. If all the groups are used, the ranks of **alendronate** \xrightarrow{T} **esophagitis**

range from 16 to 20, while they vary from 62 to 119 based on leverage. Ranks generated by UTARM lead 46 to 100 higher than these generated by OPUS_AR⁺, and 64.3 higher on average. Similarly, the ranks of nefazodone \xrightarrow{T} hepatitis are from 10 to 20, and the ranks of nefazodone \xrightarrow{T} hepatitis are from 36 to 62. UTARM sorts the association rule at least 24 and higher than OPUS_AR⁺. In addition, UTARM stably ranks the two ADRs within the top 10 for their risky age-gender groups of patients, and within the top 20 for all patients, for all the different parameter settings.

6 Discussion and Conclusions

In this paper, in order to discover the unexpected and infrequent episodes, we have proposed a new concept of Unexpected Temporal Association Rules (UTARs) and its interestingness measure, residual-leverage. Based on the new case-based exclusion and event-oriented data preparation techniques, we have developed a novel mining algorithm UTARM to discover these important UTARs. UTARM has been applied for detecting Adverse Drug Reactions (ADRs) from healthcare administrative databases. The three known ADRs, alendronate \xrightarrow{T} esophagitis, nefazodone \xrightarrow{T} hepatitis, and ACE inhibitors \xrightarrow{T} angioedema, have successfully been put in the most interesting UTARs. UTARM has also highlighted another ADR flucloxacillin \xrightarrow{T} hepatitis, of which we have not had prior knowledge until medical experts remarked on the results. In addition, the results have substantiated that UTARM can highlight these ADRs reliably with various parameter settings. Our evaluation indicates that the proposed method and measure show a promise in short-listing the possible ADRs with little prior knowledge. Considering data noise in the real-world databases, the shortlists are quite promising to help medical experts to identify ADRs effectively and efficiently. This

successful real-world application justifies the applicability of this research.

We have only concentrated on highlighting pairwise UTARs in this paper. However, the proposed concept and its interestingness measure are readily extended to detect more sophisticated UTARs, such as ADRs induced by drug interactions and disease complications. In addition, the exclusion operation opens a new direction in handling unexpectedness of unexpected patterns/temporal associations. A direct extension is to consider the other quantitative information such as dosage of drug usage during exclusion, rather than the simple set difference as we did in this paper, in order to discover undesired episodes such as diseases induced by cumulative drug toxicity. One general way for mining unexpected patterns/temporal associations is to use a data mining algorithm to find some expected ones from data directly, and then using this mined knowledge to help us discover unexpected ones. These three research directions are the subject of future work.

Acknowledgements

The authors acknowledge the Commonwealth Department of Health and Ageing, and the Queensland Department of Health for providing data for this research. We also thank Dr. Ross Sparks and Damien McAullay for constructive discussion and data preparation, respectively.

References

- [1] Australian adverse drug reaction reporting system. The Therapeutic Goods Administration (TGA), Department of Health and Ageing (DoHA), Australia.
<http://www.tga.gov.au/problem/index.htm#medicines>.

- [2] Australian adverse drug reactions bulletin. The Adverse Drug Reactions Advisory Committee (ADRAC), the Therapeutic Goods Administration (TGA), Department of Health and Ageing (DoHA), Australia. <http://www.tga.gov.au/adr/aadrb.htm>.
- [3] Australian statistics on medicines. The Drug Utilisation Sub-Committee (DUSC) of the Pharmaceutical Benefits Advisory Committee, Department of Health and Ageing (DoHA), Australia, 2000. <http://www7.health.gov.au/pbs/healthpro/pubs/asm.htm>.
- [4] R. Agrawal and R. Srikant. Mining sequential patterns. In *Proceedings of the Eleventh International Conference on Data Engineering*, pages 3–14, 1995.
- [5] D. W. Bates, D. Cullen, and al. et. Incidence of adverse drug events and potential adverse drug events. *The Journal of the American Medical Association*, 274(1):29–34, 1995.
- [6] D. W. Bates, R. S. Evans, H. Murff, P. D. Stetson, L. Pizziferri, and G. Hripcsak. Detecting adverse events using information technology. *Journal of American Medical Informatics Association*, 10(2):115–128, Mar/Apr 2003.
- [7] J. Chen, H. He, G. Williams, and H. Jin. Temporal sequence associations for rare events. In *Proceedings of PAKDD'04*, pages 235–239, May 2004.
- [8] F. R. Ernst and A. J. Grizzle. Drug-related morbidity and mortality: Updating the cost-of-illness model. *Journal of the American Pharmacists Association*, 41(2):192–199, 2001.
- [9] J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M.-C. Hsu. FreeSpan: frequent

- pattern-projected sequential pattern mining. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 355–359, 2000.
- [10] S. K. Harms and J. S. Deogun. Sequential association rule mining with time lags. *Journal of Intelligent Informatics Systems*, 22(1):7–22, 2004.
- [11] H. Jin, J. Chen, G. Williams, and H. He. A survey on temporal/streaming data mining. Technical Report CMIS 04/105, CSIRO, August 2004.
- [12] J. Lazarou, B. Pomeranz, and P. Corey. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *The Journal of the American Medical Association*, 279(15):1200–1205, Apr 1998.
- [13] C.-H. Lee, M.-S. Chen, and C.-R. Lin. Progressive partition miner: An efficient algorithm for mining general temporal association rules. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):1004–1017, 2003.
- [14] Y. Li, P. Ning, X. S. Wang, and S. Jajodia. Discovering calendar-based temporal association rules. *Data & Knowledge Engineering*, 44(2):193–218, 2003.
- [15] M. Maclure. The case-crossover design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology*, pages 144–153, Jan 1991.
- [16] H. Mannila, H. Toivonen, and A. I. Verkamo. Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, 1(3):259–289, 1997.
- [17] H. J. Murff, V. L. Patel, G. Hripcsak, and D. W. Bates. Detecting adverse events for patient safety research: a review of current methodologies. *Journal of Biomedical Informatics*, 36(1/2):131–143, 2003.

- [18] S. C. Newman. *Biostatistical Methods in Epidemiology*. John Wiley & Sons, July 2001.
- [19] J. Roberto J. Bayardo and R. Agrawal. Mining the most interesting rules. In *Proceedings of SIGKDD'99*, pages 145–154, 1999.
- [20] J. F. Roddick and M. Spiliopoulou. A survey of temporal knowledge discovery paradigms and methods. *IEEE Transactions on Knowledge and Data Engineering*, 14(4):750–767, 2002.
- [21] R. Srikant and R. Agrawal. Mining sequential patterns: generalisations and performance improvements. In *Proceedings of EDBT'96*, pages 3–17, 1996.
- [22] X. Sun, M. Orlowska, and X. Zhou. Finding event-oriented patterns in long temporal sequences. In *Proceedings of PAKDD'03*, pages 15–26, April 2003.
- [23] X. Sun, M. E. Orlowska, and X. Li. Finding negative event-oriented patterns in long temporal sequences. In *Proceedings of PAKDD'04*, pages 212–221, May 2004.
- [24] K. Wang, Y. Jiang, and L. V. Lakshmanan. Mining unexpected rules by pushing user dynamics. In *Proceedings of SIGKDD'03*, pages 246–255, 2003.
- [25] P. Wang, S. Schneeweiss, R. Glynn, H. Mogun, and J. Avorn. Use of the case-crossover design to study prolonged drug exposures and insidious outcomes. *Annals of Epidemiology*, 14:296–303, Apr 2004.
- [26] G. I. Webb. Efficient search for association rules. In *Proceedings of SIGKDD'00*, pages 99–107, 2000.
- [27] G. Williams, D. Vickers, R. Baxter, S. Hawkins, C. Kelman, R. Solon, H. He, and L. Gu.

The Queensland Linked Data Set. Technical Report CMIS 02/21, CSIRO, Canberra, 2002.

- [28] C. Zhang and S. Zhang. *Association rule mining: models and algorithms*. Springer-Verlag New York, Inc., 2002.