

Medication Use and the Risk of Newly Diagnosed Diabetes in Patients with Epilepsy: A Data Mining Application on a Healthcare Database

Sheng-Feng Sung, Ditmanson Medical Foundation Chiayi Christian Hospital, Chiayi, Taiwan & National Chung Cheng University, Minxiong, Taiwan

Pei-Ju Lee, National Chung Cheng University, Minxiong, Taiwan

Cheng-Yang Hsieh, Tainan Sin Lau Hospital and National Cheng Kung University Hospital and College of Medicine, Tainan, Taiwan

 <https://orcid.org/0000-0002-8772-4073>

Wan-Lun Zheng, National Chung Cheng University, Minxiong, Taiwan

ABSTRACT

Epilepsy is a common neurological disorder that affects millions of people worldwide. Patients with epilepsy generally require long-term antiepileptic therapy and many of them receive polypharmacy. Certain medications, including older-generation antiepileptic drugs, have been known to predispose patients to developing diabetes. Although data mining techniques have become widely used in healthcare, they have seldom been applied in this clinical problem. Here, the authors used association rule mining to discover drugs or drug combinations that may be associated with newly diagnosed diabetes. Their findings indicate in addition to the most common culprits such as phenytoin and valproic acid, prescriptions containing carbamazepine, oxcarbazepine, or lamotrigine may be related to the development of newly diagnosed diabetes. These mined rules are useful as guidance to both clinical practice and future research.

KEYWORDS

Antiepileptic Drug, Association Rules, Diabetes, Drug-Drug Interaction, Epilepsy

INTRODUCTION

Epilepsy is a common neurological disorder characterized by paroxysmal recurrence of epileptic seizures (Moshé, Perucca, Ryvlin, & Tomson, 2015). Long-term use of one or more antiepileptic drugs (AEDs) to provide optimal seizure control has been the mainstay of treatment for epilepsy (Burakgazi & French, 2016). A survey found that the mean number of AEDs consumed per each patient with epilepsy was 1.7 ± 0.8 (range 1–4) (Eyal, Rasaby, & Ekstein, 2014). In addition, the burden of comorbidities is higher in patients with epilepsy than that in the general population (Keezer, Sisodiya, & Sander, 2016). About 56% of patients with epilepsy were concomitantly treated with at least one other prescription (Eyal et al., 2014). Therefore, prescribers should carefully consider the long-term adverse effects of AEDs—especially of those older-generation AEDs with hepatic enzyme-inducing activities and high potentials for drug-drug interactions (Brodie et al., 2013).

DOI: 10.4018/JOEUC.2020040105

This article, originally published under IGI Global's copyright on April 1, 2020 will proceed with publication as an Open Access article starting on January 21, 2021 in the gold Open Access journal, Journal of Organizational and End User Computing (converted to gold Open Access January 1, 2021), and will be distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

Prior studies have demonstrated that enzyme-inducing AEDs—such as phenytoin—are associated with markers of elevated vascular risk including increased carotid intima-media thickness (Chuang et al., 2012) and a higher risk of stroke (Hsieh, Lai, Yang, & Lin, 2013). Notably, the prevalence of diabetes is higher in patients with epilepsy (Keezer et al., 2016). It is possible that metabolic side effects of AEDs directly cause diabetes—which, in turn, leads to an increased risk of vascular diseases in patients with epilepsy. For example, phenytoin may cause hyperglycemia through decreased insulin secretion and increased insulin resistance (Fathallah, Slim, Larif, Hmouda, & Ben Salem, 2015). Valproic acid may cause impaired glucose homeostasis, overweight, and metabolic syndrome (Verrotti, Manco, Agostinelli, Coppola, & Chiarelli, 2010). Furthermore, when patients with epilepsy are getting older, the association between AEDs and the increased prevalence of diabetes will become more complicated because of a higher burden of comorbidity (Johnson et al., 2018) and probably because of the cumulative burden of more co-medications. For example, patients with epilepsy have a high risk of psychiatric disorders (e.g. depression) (Keezer et al., 2016), whereas antidepressants may increase the risk of new-onset diabetes (Nguyen, Roussin, Rousseau, Montastruc, & Montastruc, 2018; Sifis & Papazisis, 2018). Therefore, the relationships between AEDs, other co-medications, and the development of diabetes are complex and remain to be elucidated.

Data mining techniques have been used to find undiscovered patterns or to build classification models in many areas including finance and healthcare. The characteristics, large volume, and complexity, of healthcare data have elicited the use of data mining modeling (Koh & Tan, 2005).

The objective of this paper was to analyze the association between drug use in patients with epilepsy and the risk of newly diagnosed diabetes using data mining techniques. By conducting knowledge mining in a huge electronic healthcare database, the research is expected to find the interaction between AEDs and diabetes in the data and to provide reference information to clinicians regarding prescription of AEDs.

The paper unfolds as follows: Section 2 provides a literature summary both of the use of data mining techniques to detect signals of adverse drug reactions and of the application of association rule mining (ARM) to medical data. Section 3 describes the research design and methods. Section 4 presents the findings after data analysis. In Section 5, the results are interpreted and discussed. Contributions to clinical practice as well as limitations and future opportunities of this study conclude the paper.

LITERATURE REVIEW

Data Mining Techniques to Detect Adverse Drug Events

Although traditional pharmacovigilance and reporting systems have been used for decades to detect adverse drug events caused by a marketed drug (Waller, 2006), adverse events of insidious onset—such as of diabetes—may be less likely to be captured by spontaneous reporting systems. Nevertheless, with the build-up of data infrastructure, data mining techniques make the hypothesis-free screening for unsuspected or latent drug-outcome associations feasible; and they help detect signals to complement pharmacovigilance surveillance (Hallas et al., 2018). For example, after mining a reporting system database and electronic medical records of three hospitals in the US, increased blood glucose levels were found to be caused by an unexpected synergistic interaction between paroxetine and pravastatin (Tatonetti et al., 2011). Another example is the detection of a higher risk of hypothyroidism in patients taking some older-generation AEDs (Lai, Yang, Lin, & Hsieh, 2013). However, data mining techniques have rarely been applied to identify potential drug combinations that will cause newly diagnosed diabetes in patients with epilepsy—a vulnerable population in which drug interactions with AEDs are major concerns.

Association Rule Mining of Medical Data

ARM, also known as market basket analysis, is a process for discovering hidden relationships between attributes in a large-scale dataset. It has been applied to large medical datasets, such as to national health insurance databases or to spontaneous reporting system databases. For example, ARM has been used to discover comorbidity patterns of attention deficit hyperactivity disorder and borderline personality disorder in a national health insurance database (Shen, Hu, & Hu, 2017; Tai & Chiu, 2009). ARM has also been commonly used for the study of drug-drug interactions by analyzing data from adverse event reporting systems (R. Cai et al., 2017; Ibrahim, Saad, Abdo, & Sharaf Eldin, 2016).

Because a large amount of raw association rules may be generated by ARM, it is vital to reduce the number of rules to a manageable size and to remove meaningless rules for ARM in order to be successful (Yoo et al., 2012). Since medical data usually have a hierarchical structure, it is natural to apply multiple-level ARM to discover interesting information. However, in order to remove uninteresting rules, the application of some measures that quantify “interestingness” or intervention by humans in the process of ARM have been proposed (Han & Fu, 1995). Various algorithms and techniques for multiple-level ARM have also been suggested (Han & Fu, 1995; Srikant & Agrawal, 1995; Vanarse & Kasar, 2017).

METHODS

Data Source

Taiwan’s National Health Insurance program is a mandatory universal health insurance program, which covers inpatient care, outpatient care, dental care, and prescription medications and enrolls virtually all its residents. This study used data from the National Health Insurance Research Database (NHIRD), which comprises all claims data from the National Health Insurance program. Because all information in the database has been de-identified for privacy protection, this study was exempted from a full review by the Institutional Review Board of Ditmanson Medical Foundation Chia-Yi Christian Hospital (CYCH-IRB No. 2018007), and informed consent was not required.

Study Patients

Figure 1 illustrates the flowchart of study patients. The authors selected adult patients who were diagnosed (between 2010 and 2011) with epilepsy—as defined by the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code 345.xx (Hsieh et al., 2013)—and who had been prescribed at least one AED. The diagnosis code of epilepsy had to be present on at least one inpatient claim or at least two outpatient claims before patients were ascertained to have epilepsy (Sung et al., 2016). Use of AEDs was identified from the prescription claims using the Anatomical Therapeutic Chemical (ATC) code N03A. The earliest date when AEDs were prescribed was designated as the index date. Patients who were diagnosed with diabetes (ICD-9-CM code 250.xx) or prescribed with any antidiabetic drug (ATC code A10) between 12 months before the index date and 3 months after the index date were excluded (Figure 2). Patients newly diagnosed with diabetes between 3 months and 12 months after the index date were assigned to the case group. Similarly, the diagnosis of diabetes was ascertained only when patients were found to have had the diagnosis code of diabetes either on at least one inpatient claim or on at least two outpatient claims—and had been prescribed with at least one antidiabetic drug. Patients who had never been diagnosed with diabetes within two years after the index date were assigned to the non-case group (Figure 2).

Association Rule Mining

ARM was first proposed by Agrawal et al (Agrawal, Imielinski, & Swami, 1993) and is one of the most common methods of data mining. As compared with other data mining techniques, it has the main advantage of having a small number of traverses when searching the hypothesis space of itemsets,

Figure 1. Flowchart of the patients in this study. AED = antiepileptic drug.

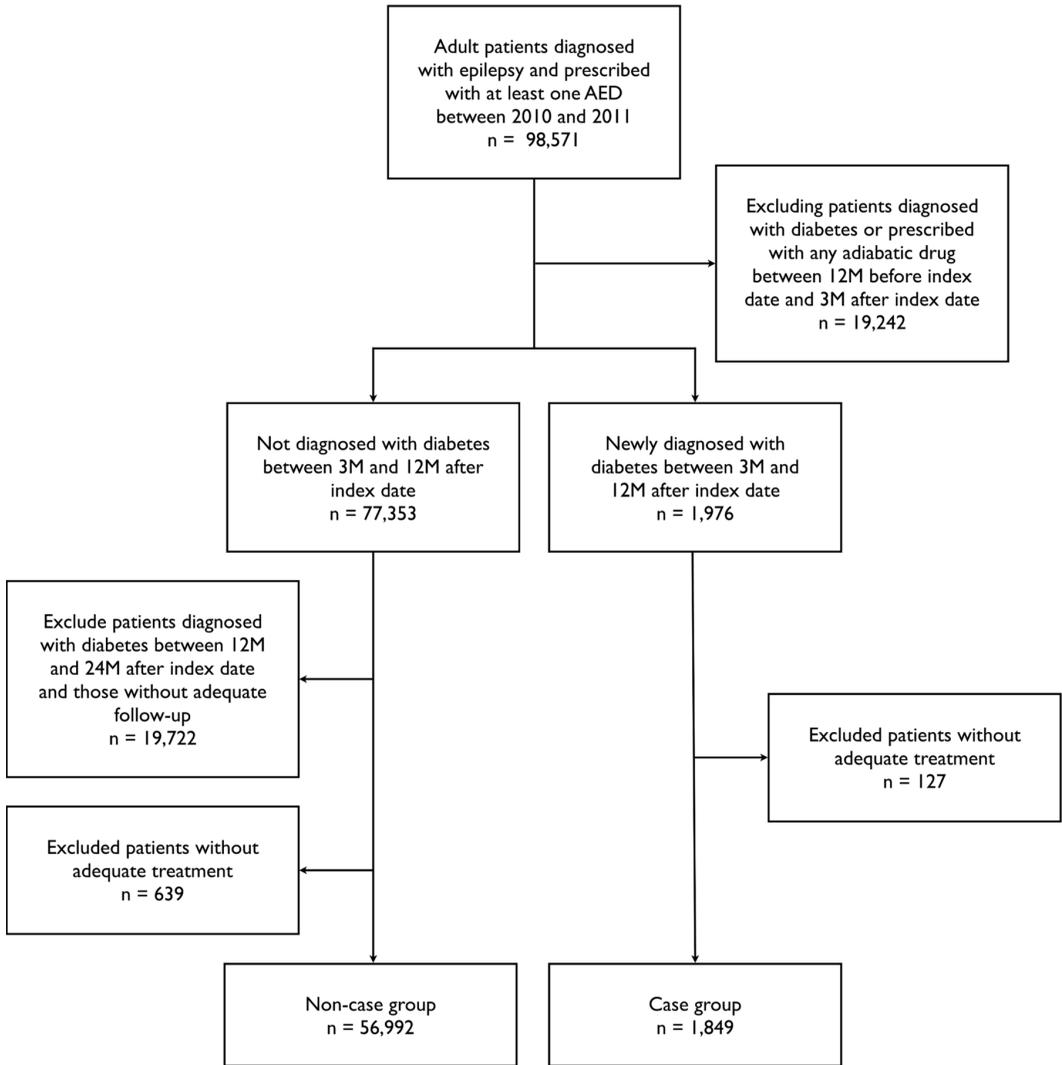
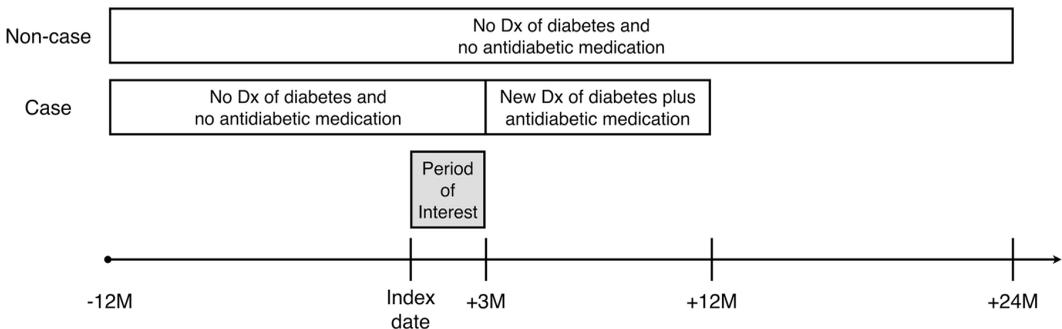


Figure 2. Schematic illustration of the study design. Dx = diagnosis.



despite the disadvantage of time complexity. The main purpose of ARM is to find the set of items of interest from a large number of transaction data. The market basket example is most often cited for ARM; the items of the basket can be represented in a binary format of whether or not a purchase was made. The relationship between products is found through transaction records made by customers and can be used to analyze customers' shopping habits, such as which products are usually purchased together. A typical example of an association rule is expressed as {milk} \rightarrow {bread}, which means people who buy milk are also likely to buy bread. The correlation analysis of products can help the store formulate marketing strategies such as using a promotion set to increase sales.

The ARM algorithm is defined as follows: if there is an association rule between two items (X and Y), it is represented by $X \rightarrow Y$ where X and Y are sets of one or more items and $X \cap Y = \emptyset$. Two conditions have to be assessed when judging the association rule $X \rightarrow Y$: one is

- **support**, which is the proportion of the amount of transaction data including ($X \cup Y$) to all transactions; and the other is
- **confidence**, which is expressed as the ratio of the number of transactions containing ($X \cup Y$) to the number of transactions containing X .

In this study, the support is the occurrence frequency of a certain drug or drugs associated with newly diagnosed diabetes in the prescription records of epileptic patients; the higher the support, the more frequently the drug(s)-diabetes itemset appears in the database. If the support and the confidence of the association rule $X \rightarrow Y$ are greater than or equal to the prespecified minimum support and minimum confidence values, then the rule is a strong association rule. The thresholds of the minimum support and minimum confidence can be determined either by the users themselves or by the experts.

Multiple-Level Association Rule Mining

In data mining applications, using different concept hierarchies to explore association rules can provide more meaningful information. The concept of multiple-level association rules was first introduced by Han and Fu in 1995 (Han & Fu, 1995). The taxonomy to construct the concept hierarchy may vary from using discrete attributes such as categories, contents, and brands (Han & Fu, 1995) to using continuous numerical attributes (Han, Cai, & Cercone, 1992). To find crossing level rules, the ML_T2L1 method proposed by Han and Fu (Han & Fu, 1995) is based on a top-down algorithm in which each concept level has its own minimum support value—and in which the minimum support of the higher level is smaller than that of the lower level. Starting from the root of the concept hierarchy to find strong rules of each level, the method sequentially finds strong rules in the next level and continues until the last level. On the other hand, the Cumulate algorithm proposed by Srikant and Agrawal (Srikant & Agrawal, 1995) finds strong rules of items at any level by considering items from the higher level in a more efficient way. Their study filtered the ancestors before adding them to itemsets, avoided ancestors that were not present, and pruned itemsets that had both the ancestor and its decedents in the same itemset.

Measures of Interestingness

Most ARM studies are dedicated to mining frequent patterns or association rules; therefore, abundant rules will be generated when the thresholds of the support and the confidence are low. However, some of the rules may not contain any useful information, while the cost of the post-processing of rules may be high. Consequently, how to balance between inducing as many rules as possible to uncover certain interesting rules and avoiding the tedious process of rule pruning will be challenging. A possible remedy is to use some measures of interestingness, which may help filter out a number of uninteresting rules.

Many objective interestingness measures have been developed in previous studies—measures such as correlation, independence, directed information ratio, etc (Brin, Motwani, & Silverstein,

1997a; Brin, Motwani, Ullman, & Tsur, 1997b; Han, Cheng, Xin, & Yan, 2007; Liu, Hsu, & Ma, 1999). Among them, the chi-square analysis is a common statistical tool—one which can be used to evaluate the dependence between the antecedent and the consequent in an association rule. For a 2×2 contingency table, a chi-square value of higher than 3.84 (95% confidence interval) will reject the independence assumption. However, the chi-square analysis is restricted to the constraints that (1) the expected values in every cell of the contingency table have to be greater than 1, and that (2) those in at least 80% of the cells need to be greater than 5 for better performance. These restrictions are not realistic in most applications of ARM (Brin et al., 1997a). Therefore, this study adopted the lift value—one of the symmetric measurement tools, also referred as interest factor—to assess the ratio of the confidence of the rule to the support of the consequent; and thus prevented a misleading result due to the high support of the consequent (Brin et al., 1997a). An interest factor of above 1 indicates a positive dependence between the antecedent and the consequent, while an interest factor below 1 shows a negative relationship. An interest factor equal to 1 means that the antecedent and the consequent are independent of each other.

Mining Association Rules Between Medication Use and Newly Diagnosed Diabetes

In this study, the authors grouped medications according to the hierarchy of the ATC classification system. It was thus a natural choice to mine multiple-level association rules. Table 1 lists the medications explored in this study. In the ATC classification system, the active substances are classified in a hierarchy with five different levels. The first level consists of 14 main anatomical/pharmacological groups. The second, third, fourth, and fifth levels are made up of therapeutic subgroups, pharmacological subgroups, chemical subgroups, and chemical substances, respectively. Through the hierarchy of medications that have been prescribed to patients, the authors could establish more comprehensive relationships between large classes and fine items by looking for itemsets based on the Apriori algorithm. For example, the rule of {N03 (antiepileptics)} → {N05AH03 (olanzapine)} indicates that the specific psycholeptic drug olanzapine may have associations with certain AEDs. On the other hand, the rule containing {N03AB02 (phenytoin)} → {N05AH03 (olanzapine)} represents a rule at a lower conceptual level and may provide more detailed information than would the previous one.

Because our primary interest was the effects of AEDs on diabetes, all drugs listed under the therapeutic subgroup N03 (antiepileptics) were investigated separately. In other words, Level 5 ATC codes were used for AEDs. For example, N03AB02 stands for phenytoin. For other study medications, both Level 2 and Level 5 ATC codes were used for multiple-level association rule mining. For example, N05 represents psycholeptics and N05AH03 stands for olanzapine. The authors categorized rules into 3 types according to what level of ATC codes were used for medications other than AEDs. When Level 5 codes were used, the mined rules were categorized as Type 1. Type 2 rules were generated when Level 2 codes were used. When both Level 2 and Level 5 codes were used, the rules were categorized as Type 3.

The period of interest was between the index date and 3 months after the index date (Figure 2). Medication use during this period was identified from the prescription claims. The defined daily dose (DDD) of each drug was obtained from the World Health Organization Collaborating Center for Drug Statistics Methodology. The authors calculated how many DDDs were dispensed per drug for each patient during this period. Following a previous study (Hung et al., 2017), patients who had not taken at least 28 cumulative DDDs of at least one AED were further excluded from the study (Figure 1) because this study focused on patients with epilepsy who were receiving antiepileptic treatment. A drug was coded as 2 if ≥28 cumulative DDDs were dispensed, as 1 if 14–27 DDDs, and as 0 if <14 DDDs.

In this study, the authors used a special type of ARM—class association rule mining (Liu, Hsu, & Ma, 1998)—to discover the relationship between AEDs, other co-medications, and the development of diabetes. ARM generally discovers all association rules that satisfy some user-specified minimum support and minimum confidence values; and the target of mining is not pre-determined. In contrast,

Table 1. ATC codes for study medications

Anatomical main groups (Level 1)	Therapeutic subgroups (Level 2)
Cardiovascular system	C01 Cardiac therapy
	C02 Antihypertensives
	C03 Diuretics
	C07 Beta blocking agents
	C10 Lipid modifying agents
Genitourinary system and sex hormones	G03 Sex hormones and modulators of the genital system
Systemic hormonal preparations, excluding sex hormones and insulins	H01 Pituitary and hypothalamic hormones and analogues
Anti-infectives for systemic use	J01 Antibacterials for systemic use
	J04 Antimycobacterials
	J05 Antivirals for systemic use
Antineoplastic and immunomodulating agents	L04 Immunosuppressants
Nervous system	N03 Antiepileptics
	N05 Psycholeptics
	N06 Psychoanaleptics
Antiparasitic products, insecticides and repellents	P01 Antiprotozoals
Respiratory system	R03 Drugs for obstructive airway diseases

ATC = Anatomical Therapeutic Chemical.

class association rules describe the co-occurrences between a set of items and a pre-defined target class. Specifically, the Apriori module in Weka 3.6 open-source machine learning software (www.cs.waikato.ac.nz/ml/weka) was used to discover interesting association rules related to the development of newly diagnosed diabetes. The “class association rules” option was enabled with the target class set to newly diagnosed diabetes. The minimum support was set to 1×10^{-4} and the minimum confidence was set to 10%. Only rules with a lift > 1 were considered as interesting.

General statistical analyses were performed using Stata 15.1 (StataCorp, College Station, Texas). Categorical variables were summarized as counts and percentages, and continuous variables as mean \pm SD. Comparisons were made using χ^2 test for categorical variables and *t*-tests for continuous variables.

RESULTS

Patient Characteristics

A total of 58,841 patients were included in the study (Figure 1). Among them, 1,849 patients developed diabetes and were assigned to the case group; whereas the remaining 56,992 patients were assigned to the non-case group. Patients in the case group were significantly older than those in the non-case group (mean age 62.0 ± 16.1 versus 46.9 ± 17.6 , $p < 0.001$) but with a similar sex distribution (female 41.8% versus 40.0%, $p = 0.124$). Table 2 lists the distribution of use of AEDs. The top three commonly used AEDs were phenytoin, valproic acid, and carbamazepine, which are all older-generation AEDs. Almost all of the AEDs were prescribed more often in the non-case group than in the case group (except for phenytoin).

Table 2. Use of antiepileptic drugs

ATC code (Level 5)	Generic name	Total N = 58,841	Case group N = 1849	Non-case group N = 56,992	P
N03AA02	phenobarbital	3,188 (5.42)	28 (1.51)	3,160 (5.54)	<0.001
N03AB02	phenytoin	23,370 (39.72)	797 (43.10)	22,573 (39.61)	0.002
N03AE01	clonazepam	13,130 (22.31)	155 (8.38)	12,975 (22.77)	<0.001
N03AF01	carbamazepine	13,464 (22.88)	163 (8.82)	13,301 (23.34)	<0.001
N03AF02	oxcarbazepine	6,509 (11.06)	104 (5.62)	6,405 (11.24)	<0.001
N03AG01	valproic acid	20,374 (34.63)	414 (22.39)	19,960 (35.02)	<0.001
N03AG04	vigabatrin	1,463 (2.49)	16 (0.87)	1,447 (2.54)	<0.001
N03AG06	tiagabine	185 (0.31)	3 (0.16)	182 (0.32)	0.235
N03AX09	lamotrigine	5,400 (9.18)	62 (3.35)	5,338 (9.37)	<0.001
N03AX11	topiramate	6,749 (11.47)	83 (4.49)	6,666 (11.70)	<0.001
N03AX12	gabapentin	3,866 (6.57)	85 (4.60)	3,781 (6.63)	0.001
N03AX14	levetiracetam	10,303 (17.51)	138 (7.46)	10,165 (17.84)	<0.001
N03AX16	pregabalin	405 (0.69)	1 (0.05)	404 (0.71)	0.001

Data are listed as numbers (percentages)

Results of Association Rule Mining

Table 3 gives the draft results of ARM. A total of 52 rules were generated using the specified minimum support ($1 \times E-04$) and confidence (10%) values and lift threshold (>1). Of them, 4, 11, and 37 rules belonged to the Type 1, Type 2, and Type 3 rule, respectively. Because this study focused on the effects of AEDs on the development of newly diagnosed diabetes, the authors only considered rules that contained at least one AED on the left-hand side of the rule. To avoid redundancy, only the Level 5 ATC code was retained when both a Level 5 code and its corresponding Level 2 code coexisted. For example, the rule $\{N03AB02=2, R03=1, R03CC12=1\} \rightarrow \{\text{diabetes}\}$ was simplified to $\{N03AB02=2, R03CC12=1\} \rightarrow \{\text{diabetes}\}$. Eleven rules which contained AEDs—including phenytoin, valproic acid, carbamazepine, oxcarbazepine, and lamotrigine—were identified (Table 4).

DISCUSSION AND IMPLICATIONS

Summary of Study Findings

In this large nationwide cohort of patients with epilepsy starting AED treatment, the authors discovered 52 association rules regarding prescriptions that were potentially associated with the development of newly diagnosed diabetes within one year after starting AED treatment. Among these rules, eleven rules included an AED with or without other co-medication(s). In addition to the commonly reported culprits such as phenytoin and valproic acid, prescriptions containing carbamazepine, oxcarbazepine, or lamotrigine may be related to the development of newly diagnosed diabetes.

Known Drugs Causing Diabetes

As reported in the literature (Anyanwagu, Idris, & Donnelly, 2016; Fathallah et al., 2015), the authors found that several classes of cardiovascular therapeutics—including diuretics (ATC code C03), beta blocking agents (C07), and lipid-modifying agents (C10) (in particular, atorvastatin (C10AA05))—were associated with the development of diabetes. Other well-known culprit drug classes—such as

Table 3. Draft results of association rule mining for medication use and newly diagnosed diabetes ({medication(s)} → {diabetes})

Rules	N	Support (E-04)	Confidence	Lift
Type 1				
{C01DA14=2, N03AG01=1}	6	1.02	0.14	4.55
{N03AF01=1}	52	8.84	0.11	3.51
{C01DA08=1}	11	1.87	0.11	3.40
{C01AA05=1}	12	2.04	0.10	3.26
Type 2				
{C03=2, N03AX09=1}	6	1.02	0.19	6.16
{C01=2, N03AF02=1}	6	1.02	0.15	4.90
{C03=2, N03AF02=1}	7	1.19	0.15	4.74
{C07=1, N06=1}	12	2.04	0.14	4.49
{N03AB02=1, N05=1}	12	2.04	0.14	4.39
{C01=2, C10=1}	7	1.19	0.13	3.98
{N03AG01=1, N05=1}	13	2.21	0.12	3.66
{N03AF01=1}	52	8.84	0.11	3.51
{C01=2, C07=1, N03AB02=2}	9	1.53	0.11	3.49
{N03AG01=1, N06=1}	9	1.53	0.10	3.33
{C07=1, N05=1}	11	1.87	0.10	3.27
Type 3				
{C01=2, C10=1, C10AA05=1}	6	1.02	0.21	6.82
{C03=2, N03AX09=1}	6	1.02	0.19	6.16
{C01=1, C01DA08=1}	9	1.53	0.16	5.02
{C01=2, N03AF02=1}	6	1.02	0.15	4.90
{C07=1, C07AG02=1, N03AB02=2}	9	1.53	0.15	4.77
{C03=2, N03AF02=1}	7	1.19	0.15	4.74
{N05CD03=2, N06=1, N06AX05=1}	6	1.02	0.15	4.66
{N05=2, N05CD03=2, N06=1, N06AX05=1}	6	1.02	0.15	4.66
{C07=1, C07AG02=1}	21	3.57	0.14	4.61
{C01=1, C01DX16=1}	8	1.36	0.14	4.55
{C01=1, C01AA05=1}	11	1.87	0.14	4.55
{C01AA05=2, C07=1}	7	1.19	0.14	4.55
{C01DA14=2, N03AG01=1}	6	1.02	0.14	4.55
{N03AB02=2, R03=1, R03CC12=1}	6	1.02	0.14	4.55
{C01=2, C01AA05=2, C07=1}	7	1.19	0.14	4.55
{C01=2, C01DA14=2, N03AG01=1}	6	1.02	0.14	4.55
{C07=1, N06=1}	12	2.04	0.14	4.49
{C07AB07=1, N05=1}	6	1.02	0.14	4.44
{N03AB02=1, N05=1}	12	2.04	0.14	4.39
{C01=2, C07=1, C07AG02=1}	6	1.02	0.13	4.24
{C01=2, C10AA05=1}	6	1.02	0.13	4.15
{C01=2, C10=1}	7	1.19	0.13	3.98
{C07=1, C07AG01=1}	8	1.36	0.12	3.74
{N05=1, N05BX01=1, R03=2}	6	1.02	0.12	3.74
{N03AG01=1, N05=1}	13	2.21	0.12	3.66
{N05CD04=1, N06=1}	7	1.19	0.11	3.59
{N03AF01=1}	52	8.84	0.11	3.51
{C01=2, C07=1, N03AB02=2}	9	1.53	0.11	3.49
{C01DA08=1}	11	1.87	0.11	3.40
{N03AG01=1, N06=1}	9	1.53	0.10	3.33
{C07=1, C07AB07=1}	41	6.97	0.10	3.29
{C07=1, N05=1}	11	1.87	0.10	3.27
{C01AA05=1}	12	2.04	0.10	3.26
{C01=1, C03CA01=2}	10	1.70	0.10	3.25
{C01=1, C03=2, C03CA01=2}	10	1.70	0.10	3.25
{C01=2, C07=1, C07AB07=1}	8	1.36	0.10	3.22
{N05=1, N05CF02=1, N06=2}	8	1.36	0.10	3.22

Table 4. Results of association rule mining for medication use containing antiepileptic drugs and newly diagnosed diabetes ({medication(s)} → {diabetes})

Rules	N	Support (E-04)	Confidence	Lift
{phenytoin=2, carvedilol=1}	9	1.53	0.15	4.77
{phenytoin=2, bambuterol=1}	6	1.02	0.14	4.55
{phenytoin=1, PSYCHOLEPTICS=1}	12	2.04	0.14	4.39
{phenytoin=2, CARDIAC THERAPY= 2, BETA BLOCKING AGENTS=1}	9	1.53	0.11	3.49
{valproic acid=1, isosorbide mononitrate=2}	6	1.02	0.14	4.55
{valproic acid=1, PSYCHOLEPTICS=1}	13	2.21	0.12	3.66
{valproic acid=1, PSYCHOANALEPTICS=1}	9	1.53	0.10	3.33
{carbamazepine=1}	52	8.84	0.11	3.51
{oxcarbazepine=1, CARDIAC THERAPY=2}	6	1.02	0.15	4.90
{oxcarbazepine=1, DIURETICS=2}	7	1.19	0.15	4.74
{lamotrigine=1, DIURETICS=2}	6	1.02	0.19	6.16

psycholeptics (N05), psychoanaleptics, (N06) or drugs for obstructive airway diseases (R03)—were also reproducibly detected. To the best of our knowledge, ARM has been rarely applied to detect adverse drug effects such as newly diagnosed diabetes. Our results confirmed the potential of ARM as an effective method to detect signals in pharmacovigilance studies.

Among the 11 association rules regarding prescription of AEDs, phenytoin and valproic acid—as expected—were included in seven of them. For decades, phenytoin has been known to induce hyperglycemia. The potential mechanisms may involve impaired insulin release (Fariss & Lutchter, 1971) and increased insulin insensitivity through the induction of a post-binding defect in insulin action (al-Rubeaan & Ryan, 1991). On the other hand, valproic acid may cause overweightness and impaired glucose homeostasis and thus may increase the risk of diabetes (Verrotti et al., 2010). Notably, the burden of comorbidities such as heart disease, depression, and dementia are up to eight times higher in patients with epilepsy than in the general population (Keezer et al., 2016). Comedications for treating these comorbidities—including beta blocking agents, psycholeptics, and psychoanaleptics—were found in the association rules containing phenytoin or valproic acid. These rules may be of great clinical significance in guiding physicians to monitor blood glucose closely when such drug combinations are used.

Previously Unknown Associations

The authors identified several AEDs other than phenytoin and valproic acid in the remaining four association rules. For example, carbamazepine alone was associated with newly diagnosed diabetes. Carbamazepine is the first choice of epilepsy with focal onset in adults (Burakgazi & French, 2016) and is widely used locally (Lai et al., 2016) because it is effective and relatively inexpensive. Its high frequency of prescription may explain the high value of support (8.84 E-04) for this rule. Similar to valproic acid, carbamazepine may be associated with metabolic syndrome and insulin resistance (Güler, Güneş, of, 2016, 2016). Another study also reported that carbamazepine was associated with an even higher risk of elevated fasting serum glucose as compared to valproic acid (Rakitin, Kōks, & Haldre, 2016).

Besides, the authors also found that drug combinations of oxcarbazepine with cardiac therapy or diuretics, and lamotrigine with diuretics were associated with newly diagnosed diabetes. Although the

link between diuretics and newly diagnosed diabetes is well known (Anyanwagu et al., 2016; Fathallah et al., 2015), the risk of developing diabetes is less described for oxcarbazepine and lamotrigine. So far, diabetes has only accounted for a small proportion of reported adverse drug reactions to these two drugs, according to online reports. Readers may refer to “Lamotrigine and Diabetes mellitus – from FDA reports” (<https://www.ehealthme.com/ds/lamotrigine/diabetes-mellitus/>) and “Trileptal and Type 2 diabetes – from FDA reports” (<https://www.ehealthme.com/ds/trileptal/type-2-diabetes/>). The authors are not aware of any published studies suggesting the association between oxcarbazepine or lamotrigine and newly diagnosed diabetes. Further studies are warranted to confirm or refute these association rules, because either oxcarbazepine or lamotrigine is used as the first-line monotherapy in adult patients with epilepsy (Faught, Helmers, Thurman, Kim, & Kalilani, 2018; Lai et al., 2016).

Application of Association Rule Mining to Detect Adverse Drug Effects

Conventional pharmacoepidemiological studies, such as those using the cohort study design, may provide a straightforward way to evaluate the association between drug exposures and adverse outcomes. However, such study designs may be hard to perform when drug exposures are complex and time varying (Pottegård, Friis, Stürmer, Hallas, & Bahmanyar, 2018), as is seen in the case of patients with epilepsy. In contrast, ARM may herein serve as an alternative way to find potential drugs or drug combinations which may have an association with the risk of developing diabetes. In our view, ARM may not and should not replace conventional pharmacoepidemiological studies in the confirmation of adverse drug reactions. Nevertheless, ARM, as a complementary tool, can facilitate the screening and exploration of the relationships between AEDs, various drug classes, and diabetes—thus providing useful signals or hypotheses for future confirmatory prospective studies.

Suggestion

This study verified the findings of previous studies that certain prescribed medications are associated with newly diagnosed diabetes such as older-generation AEDs, several classes of cardiovascular therapeutics, psycholeptics, and psychoanaleptics. Moreover, several AEDs—including carbamazepine, oxcarbazepine, and lamotrigine—were revealed to be associated with newly diagnosed diabetes in this work. In addition to monitoring the adverse effects of a single drug, the detection of adverse effects of drug combinations has become an important issue in pharmacovigilance. Therefore, this study provides an example of how data mining techniques can contribute to pharmacovigilance and clinical practice.

Limitations

Certain limitations were present in this study.

Firstly, the association between medication use and newly diagnosed diabetes is not necessarily a causal one. It may be confounded by other factors that are associated with the use of certain medication and the development of diabetes. For example, shared risk factors (e.g., genetics) may predispose patients to developing epilepsy and mood disorders simultaneously, thereby leading to the co-prescription of valproic acid and psycholeptics or psychoanaleptics; while sedentary lifestyle and unhealthy diet habit due to mood disorders may contribute to the development of diabetes.

Secondly, the possibility of a reverse causality cannot be ruled out. In other words, a patient with covert diabetes may develop an epileptic seizure as the first manifestation of diabetes (Omar, El-Khabiry, & Vaughan, 2012) and may thus be prescribed AEDs. A presumption of a causal association is that the exposure to the risk factor (medication use) occurs first and is later followed by the occurrence of the disease (diabetes). Hence, the use of a three-month lag time period between medication use and newly diagnosed diabetes in the study design may help ameliorate the risk of reverse causation (Pottegård et al., 2018).

Thirdly, the mined associations could be artifactual as a result of surveillance bias. For example, the prescription of phenytoin, valproic acid, or carbamazepine often requires repeated blood tests for

therapeutic drug monitoring. Patients prescribed with these drugs were likely to do blood glucose tests at the same time and thus may receive a diagnosis of diabetes earlier than those prescribed with other antiepileptic drugs which do not need therapeutic drug monitoring. Similar surveillance bias might also happen in patients prescribed with cardiac therapy, isosorbide mononitrate, and beta blocking agents for cardiovascular diseases because aggressive investigation of vascular risk factors is indicated for such patients.

Finally, the authors were unable to adjust for several important confounders such as age. Consequently, the authors could not exclude the possibility that some association rules hold only because certain antiepileptic drugs or drug combinations were more likely to be prescribed to elderly patients whose risk of diabetes is relatively high. However, new users of lamotrigine or oxcarbazepine were younger than those of other AEDs (Lai et al., 2016). Therefore, the association rules containing either of the two drugs might be free from the potentially confounding influence of age.

CONCLUSION

This study used ARM to explore the relationships between medication use and newly diagnosed diabetes in adult patients with epilepsy—and discovered several rules that may be useful for physicians and researchers. In accordance with previous literature, phenytoin and valproic acid—when used in combination with other drug classes such as beta blocking agents, psycholeptics, and psychoanaleptics—accounted for more than half of the mined association rules. In addition, the authors found rules containing carbamazepine, oxcarbazepine, or lamotrigine that had not been reported in previous literature. Such rules may serve as hypotheses to be tested in future studies. Furthermore, our work may support the use of ARM in the field of pharmacoepidemiology to both detect signals and generate hypotheses between drug exposures and adverse events.

ACKNOWLEDGMENT

This study is based in part on data from the National Health Insurance Research Database provided by the Bureau of National Health Insurance, Department of Health and managed by National Health Research Institutes. The interpretation and conclusions contained herein do not represent those of Bureau of National Health Insurance, Department of Health, or National Health Research Institutes. The authors would like to thank Dr. Chin-Wei Huang for his role in acquiring the data. This research was supported in part by the Ministry of Science and Technology (Grant Number MOST 107-2314-B-705-001 and MOST 107-2410-H-194-053).

REFERENCES

- Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large database. In *SIGMOD '93 Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data* (pp. 207–216). Association for Computing Machinery.
- al-Rubeaan, K., & Ryan, E. A. (1991). Phenytoin-induced insulin insensitivity. *Diabetic Medicine*, 8(10), 968–970. doi:10.1111/j.1464-5491.1991.tb01539.x PMID:1838051
- Anyanwagu, U., Idris, I., & Donnelly, R. (2016). Drug-induced diabetes mellitus: Evidence for statins and other drugs affecting glucose metabolism. *Clinical Pharmacology and Therapeutics*, 99(4), 390–400. doi:10.1002/cpt.274 PMID:26440603
- Brin, S., Motwani, R., & Silverstein, C. (1997a). Beyond market baskets: Generalizing association rules to correlations. In *SIGMOD '97 Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data* (pp. 265–276). Association for Computing Machinery.
- Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997b). Dynamic itemset counting and implication rules for market basket data. In *SIGMOD '97 Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data* (pp. 255–264). Association for Computing Machinery.
- Brodie, M. J., Mintzer, S., Pack, A. M., Gidal, B. E., Vecht, C. J., & Schmidt, D. (2013). Enzyme induction with antiepileptic drugs: Cause for concern? *Epilepsia*, 54(1), 11–27. doi:10.1111/j.1528-1167.2012.03671.x PMID:23016553
- Burakgazi, E., & French, J. A. (2016). Treatment of epilepsy in adults. *Epileptic Disorders: International Epilepsy Journal with Videotape*, 18(3), 228–239. doi:10.1684/epd.2016.0836 PMID:27435035
- Cai, R., Liu, M., Hu, Y., Melton, B. L., Matheny, M. E., Xu, H., & Waitman, L. R. et al. (2017). Identification of adverse drug-drug interactions through causal association rule discovery from spontaneous adverse event reports. *Artificial Intelligence in Medicine*, 76, 7–15. doi:10.1016/j.artmed.2017.01.004 PMID:28363289
- Chuang, Y.-C., Chuang, H.-Y., Lin, T.-K., Chang, C.-C., Lu, C.-H., Chang, W.-N., & Chan, S. H. H. et al. (2012). Effects of long-term antiepileptic drug monotherapy on vascular risk factors and atherosclerosis. *Epilepsia*, 53(1), 120–128. doi:10.1111/j.1528-1167.2011.03316.x PMID:22085257
- Eyal, S., Rasaby, S., & Ekstein, D. (2014). Concomitant therapy in people with epilepsy: Potential drug-drug interactions and patient awareness. *Epilepsy & Behavior*, 31, 369–376. doi:10.1016/j.yebeh.2013.09.041 PMID:24211058
- Fariss, B. L., & Lutcher, C. L. (1971). Diphenylhydantoin-induced hyperglycemia and impaired insulin release. Effect of dosage. *Diabetes*, 20(3), 177–181. doi:10.2337/diab.20.3.177 PMID:5552685
- Fathallah, N., Slim, R., Larif, S., Hmouda, H., & Ben Salem, C. (2015). Drug-induced hyperglycaemia and diabetes. *Drug Safety*, 38(12), 1153–1168. doi:10.1007/s40264-015-0339-z PMID:26370106
- Faught, E., Helters, S., Thurman, D., Kim, H., & Kalilani, L. (2018). Patient characteristics and treatment patterns in patients with newly diagnosed epilepsy: A US database analysis. *Epilepsy & Behavior: E&B*, 85, 37–44. doi:10.1016/j.yebeh.2018.05.019
- Güler, S. K., & Güneş, N., of, B. Ç. E. J., 2016. (2016). Development of insulin resistance in patients with epilepsy during valproate and carbamazepine monotherapy. *Epilepsi*, 22(3), 102–107. doi:10.5505/epilepsi.2016.88700
- Hallas, J., Wang, S. V., Gagne, J. J., Schneeweiss, S., Pratt, N., & Pottegård, A. (2018). Hypothesis-free screening of large administrative databases for unsuspected drug-outcome associations. *European Journal of Epidemiology*, 33(6), 545–555. doi:10.1007/s10654-018-0386-8 PMID:29605890
- Han, J., Cai, Y., & Cercone, N. (1992). Knowledge discovery in databases: An attribute-oriented approach. In *VLDB '92 Proceedings of the 18th International Conference on Very Large Data Bases* (pp. 547–559). Association for Computing Machinery.
- Han, J., Cheng, H., Xin, D., & Yan, X. (2007). Frequent pattern mining: Current status and future directions. *Data Mining and Knowledge Discovery*, 15(1), 55–86. doi:10.1007/s10618-006-0059-1

Han, J., & Fu, Y. (1995). Discovery of multiple-level association rules from large databases. In *VLDB '95 Proceedings of the 21th International Conference on Very Large Data Bases* (pp. 420–431). Association for Computing Machinery.

Hsieh, C.-Y., Lai, E. C.-C., Yang, Y.-H. K., & Lin, S.-J. (2013). Comparative stroke risk of antiepileptic drugs in patients with epilepsy. *Epilepsia*, *54*(1), 172–180. doi:10.1111/j.1528-1167.2012.03693.x PMID:23030457

Hung, M.-S., Chen, I.-C., Lee, C.-P., Huang, R.-J., Chen, P.-C., Tsai, Y.-H., & Yang, Y.-H. (2017). Statin improves survival in patients with EGFR-TKI lung cancer: A nationwide population-based study. *PLoS One*, *12*(2), e0171137. doi:10.1371/journal.pone.0171137 PMID:28158206

Ibrahim, H., Saad, A., Abdo, A., & Sharaf Eldin, A. (2016). Mining association patterns of drug-interactions using post marketing FDA's spontaneous reporting data. *Journal of Biomedical Informatics*, *60*, 294–308. doi:10.1016/j.jbi.2016.02.009 PMID:26903152

Johnson, E. L., Krauss, G. L., Lee, A. K., Schneider, A. L. C., Dearborn, J. L., Kucharska-Newton, A. M., & Gottesman, R. F. et al. (2018). Association between midlife risk factors and late-onset epilepsy: Results from the Atherosclerosis Risk in Communities Study. *JAMA Neurology*, *75*(11), 1375–1382. doi:10.1001/jamaneurol.2018.1935 PMID:30039175

Keezer, M. R., Sisodiya, S. M., & Sander, J. W. (2016). Comorbidities of epilepsy: Current concepts and future perspectives. *Lancet Neurology*, *15*(1), 106–115. doi:10.1016/S1474-4422(15)00225-2 PMID:26549780

Koh, H. C., & Tan, G. (2005). Data mining applications in healthcare. *Journal of Healthcare Information Management*, *19*(2), 64–72. PMID:15869215

Lai, E. C.-C., Hsieh, C.-Y., Su, C.-C., Yang, Y.-H. K., Huang, C.-W., Lin, S.-J., & Setoguchi, S. (2016). Comparative persistence of antiepileptic drugs in patients with epilepsy: A STROBE-compliant retrospective cohort study. *Medicine*, *95*(35), e4481. doi:10.1097/MD.0000000000004481 PMID:27583857

Lai, E. C.-C., Yang, Y.-H. K., Lin, S.-J., & Hsieh, C.-Y. (2013). Use of antiepileptic drugs and risk of hypothyroidism. *Pharmacoepidemiology and Drug Safety*, *22*(10), 1071–1079. doi:10.1002/pds.3498 PMID:23946049

Liu, B., Hsu, W., & Ma, Y. (1998). Integrating classification and association rule mining. In *KDD'98 Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining* (pp. 80–86). Association for Computing Machinery.

Liu, B., Hsu, W., & Ma, Y. (1999). Pruning and summarizing the discovered associations. In *KDD '99 Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 125–134). Association for Computing Machinery.

Moshé, S. L., Perucca, E., Ryvlin, P., & Tomson, T. (2015). Epilepsy: New advances. *Lancet*, *385*(9971), 884–898. doi:10.1016/S0140-6736(14)60456-6 PMID:25260236

Nguyen, T. T. H., Roussin, A., Rousseau, V., Montastruc, J.-L., & Montastruc, F. (2018). Role of serotonin transporter in antidepressant-induced diabetes mellitus: A pharmacoepidemiological-pharmacodynamic study in VigiBase®. *Drug Safety*, *314*(3), 1818. doi:10.1007/s40264-018-0693-8 PMID:29956218

Omar, H. R., El-Khabiry, E., & Vaughan, S. (2012). Seizure as the first presentation of diabetes mellitus. *Therapeutic Advances in Endocrinology and Metabolism*, *3*(5), 175–177. doi:10.1177/2042018812459879 PMID:23185689

Pottegård, A., Friis, S., Stürmer, T., Hallas, J., & Bahmanyar, S. (2018). Considerations for pharmacoepidemiological studies of drug-cancer associations. *Basic & Clinical Pharmacology & Toxicology*, *122*(5), 451–459. doi:10.1111/bcpt.12946 PMID:29265740

Rakitin, A., Köks, S., & Haldre, S. (2016). Metabolic syndrome and anticonvulsants: A comparative study of valproic acid and carbamazepine. *Seizure: the Journal of the British Epilepsy Association*, *38*, 11–16. doi:10.1016/j.seizure.2016.03.008 PMID:27061880

Shen, C.-C., Hu, L.-Y., & Hu, Y.-H. (2017). Comorbidity study of borderline personality disorder: Applying association rule mining to the Taiwan national health insurance research database. *BMC Medical Informatics and Decision Making*, *17*(1), 8. doi:10.1186/s12911-016-0405-1 PMID:28077135

- Siafis, S., & Papazisis, G. (2018). Detecting a potential safety signal of antidepressants and type 2 diabetes: A pharmacovigilance-pharmacodynamic study. *British Journal of Clinical Pharmacology*, *84*(10), 2405–2414. doi:10.1111/bcp.13699 PMID:29953643
- Srikant, R., & Agrawal, R. (1995). Mining generalized association rules. In *VLDB '95 Proceedings of the 21th International Conference on Very Large Data Bases* (pp. 407–419). Association for Computing Machinery.
- Sung, S.-F., Hsieh, C.-Y., Lin, H.-J., Chen, Y.-W., Yang, Y.-H. K., & Li, C.-Y. (2016). Validation of algorithms to identify stroke risk factors in patients with acute ischemic stroke, transient ischemic attack, or intracerebral hemorrhage in an administrative claims database. *International Journal of Cardiology*, *215*, 277–282. doi:10.1016/j.ijcard.2016.04.069 PMID:27128546
- Tai, Y.-M., & Chiu, H.-W. (2009). Comorbidity study of ADHD: Applying association rule mining (ARM) to National Health Insurance Database of Taiwan. *International Journal of Medical Informatics*, *78*(12), e75–e83. doi:10.1016/j.ijmedinf.2009.09.005 PMID:19853501
- Tatonetti, N. P., Denny, J. C., Murphy, S. N., Fernald, G. H., Krishnan, G., Castro, V., & Altman, R. B. et al. (2011). Detecting drug interactions from adverse-event reports: Interaction between paroxetine and pravastatin increases blood glucose levels. *Clinical Pharmacology and Therapeutics*, *90*(1), 133–142. doi:10.1038/clpt.2011.83 PMID:21613990
- Vanarse, M., & Kasar, S. (2017). Multilevel association rule mining for large datasets: A review. *International Journal of Advanced Research in Computer Science*, *8*(8), 583–586. doi:10.26483/ijarcs.v8i8.4793
- Verrotti, A., Manco, R., Agostinelli, S., Coppola, G., & Chiarelli, F. (2010). The metabolic syndrome in overweight epileptic patients treated with valproic acid. *Epilepsia*, *51*(2), 268–273. doi:10.1111/j.1528-1167.2009.02206.x PMID:19682024
- Waller, P. C. (2006). Making the most of spontaneous adverse drug reaction reporting. *Basic & Clinical Pharmacology & Toxicology*, *98*(3), 320–323. doi:10.1111/j.1742-7843.2006.pto_286.x PMID:16611209
- Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J.-F., & Hua, L. (2012). Data mining in healthcare and biomedicine: A survey of the literature. *Journal of Medical Systems*, *36*(4), 2431–2448. doi:10.1007/s10916-011-9710-5 PMID:21537851

Sheng-Feng Sung received his MD from the National Taiwan University in 1991. He was trained at the National Taiwan University Hospital for residency. He is an attending neurologist and the head of the division of neurology at the Ditmanson Medical Foundation Chiayi Christian Hospital. He received his MS in Information Management from the National Chung Cheng University in 2015. Currently, he is also a PhD student in Information Management at the National Chung Cheng University. His research interests include clinical neurology, stroke, medical informatics, and data mining. He has published widely in scholarly journals in clinical medicine, and healthcare sciences.

Pei-Ju Lee is currently an associate professor of Department of Information Management at National Chung Cheng University, Taiwan. She received a PhD degree in Information Sciences from University of Pittsburgh in 2015. Her current research interests include information fusion, data mining, database management, human-computer interaction, and human factor.

Cheng-Yang Hsieh received his MD from the National Cheng Kung University in 2003. He was trained at the National Cheng Kung University Hospital for residency from 2005 to 2009. He is now an attending neurologist at the Tainan Sin Lau Hospital. He also received his PhD in Clinical Pharmacy from the National Cheng Kung University in 2014. Currently, he is also an adjunct assistant professor at the National Cheng Kung University. His research interests include clinical neurology, stroke, and pharmacoepidemiology. He has published widely in scholarly journals in clinical medicine, and healthcare sciences.

Wan-Lun Zheng received her MS degree in Graduate Institute of Healthcare Information Management from National Chung Cheng University of Taiwan in 2018. Her research interests include data mining, information retrieval, and text mining.