DISCOVERING OF MULTI-DIMENTIONAL ASSOCIATION RULE IN SCHIZOPHRENIA PATIENS WITH SUBSTANCE USE

WICHIAN BOONYAPRAPA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(INFORMATION TECHNOLOGY MANAGEMENT)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2015

Thesis
entitled

DISCOVERING OF MULTI-DIMENTIONAL ASSOCIATION RULE IN SCHIZOPHRENIA PATIENTS WITH SUBSTANCE USE

Wichian Boonyaprapa
.....
Mr. Wichian Boonyaprapa
Candidate

.....
Lect. Sotarat Thammaboosadee,
Ph.D. (Information Technology)
Major advisor

.....
Asst. Prof. Supaporn Kiattisin,
Ph.D. (Electrical and Computer Engineering)
Co-advisor

Adisorn Leelasantitham
.....
Asst. Prof. Adisorn Leelasantitham,
Ph.D. (Electrical Engineering)
Co-advisor

.....
Prof. Patcharee Lertrit,
M.D., Ph.D. (Biochemistry)
Dean
Faculty of Graduate Studies
Mahidol University

.....
Asst. Prof. Supaporn Kiattisin,
Ph.D. (Electrical and Computer Engineering)
Program Director
Master of Science Program in Information Technology Management
Faculty of Engineering,
Mahidol University

Thesis
entitled

DISCOVERING OF MULTI-DIMENTIONAL ASSOCIATION RULE IN SCHIZOPHRENIA PATIENTS WITH SUBSTANCE USE

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science
(Information Technology Management)
on
December 24, 2015

Wichian Boonyaprapa
.....
Mr. Wichian Boonyaprapa
Candidate

.....
Asst. Prof. Nantawat Sitdhiraksa,
MD., Ph.D. (Psychiatry)
Chair

.....
Asst. Prof. Supaporn Kiattisin,
Ph.D. (Electrical and Computer Engineering)
Member

Adisorn Leelasantitham
.....
Asst. Prof. Adisorn Leelasantitham,
Ph.D. (Electrical Engineering)
Member

Sotarat Th.
.....
Lect. Sotarat Thammaboosadee,
Ph.D. (Information Technology)
Member

.....
Prof. Somkiat Wattanasirichaigoon,
M.D., FRCST (General Surgery)
Member

.....
Prof. Patcharee Lertrit,
M.D., Ph.D. (Biochemistry)
Dean
Faculty of Graduate Studies
Mahidol University

.....
Asst.Prof. Jackrit Suthakorn,
Ph.D. (Robotics)
Dean
Faculty of Engineering
Mahidol University

ACKNOWLEDGEMENTS

I would also sincerely thank to Siriraj Development Scholarship, Faculty of Medicine Siriraj Hospital for the full scholarship, and this thesis is partially supported by Graduate Studies of Mahidol University Alumni Association.

I would also like to deeply grateful to the success of this thesis can be succeeded by the attentive support from my major advisor Lecturer. Dr. Sotarat Thammaboosadee for his invaluable advice and his kindness in writing format proofing and provide recommendations and monitor the progress of the thesis with attention all along.

I would also like to extend my sincere thanks to Assist. Prof. Nantawat Sitdhiraksa, MD, Lecturer. Dr. Woraphat Ratthaapha, MD, residency and medical student from the Department of Psychiatry, Faculty of Medicine of Siriraj Hospital, Mahidol University for their valuable advice for this thesis. Their ideas are so useful that thesis becomes more helpful for traditional Psychiatric personnel.

Furthermore, I am grateful to Research Ethics Committee of the Siriraj Hospital, Siriraj Hospital, Department of Siriraj Information Technology, Department Record Division Siriraj Hospital and Ms. Wandee Wansrisuthon for advice assistance and facilitate the collection of data as well for gave me useful information in order to study this thesis.

Finally, I am very grateful to my family members for attending to my study care, love, and encouragement. For this successfulness, I dedicate to my parents and all the teachers that are of inspiration to push me to future success. I would like to thank to all of those who I have not listed above.

Wichian Boonyaprapa

DISCOVERING OF MULTI-DIMENTIONAL ASSOCIATION RULE IN SCHIZOPHRENIA PATIENTS WITH SUBSTANCE USE

WICHIAN BOONYAPRAPA 5636643 EGIT/M

M.Sc.(INFORMATION TECHNOLOGY MANAGEMENT)

THESIS ADVISORY COMMITTEE : SOTARAT THAMMABOOSADEE, Ph.D., SUPAPORN KIATTISIN, Ph.D., ADISORN LEELASANTITHAM, Ph.D.

ABSTRACT

This thesis aims to study multi-dimensional association rules and data visualization for the schizophrenia outpatient data, which are recorded in ICD-10 format. The data comprised of 4 main dimensions: psychoactive substance use, duration of substance use, patient demographic data, and schizophrenia status. A total of 24 datasets were extracted from those main dimensions, which specified schizophrenia as the target. The processed datasets discovered the associations using the Frequent Patterns-Growth algorithm (FP-Growth), which is an extension of the traditionally used one, the Apriori algorithm, to determine the rules representing the association of schizophrenia with the other factors. The results revealed the association rules between schizophrenia and substance use had 5 main features: 1) Thai men 35-44 years old, 2) Bangkok residents, 3) nicotine use history, 4) amphetamine use history in a period of 7-12 months, and 5) diagnosis of substance dependence. All of the selected association rules had an acceptable confidence level over 0.90. This could confirm that the multi-dimensional association rules driven by the FP-Growth algorithm could be an appropriate technique to demonstrate the relationship pattern of data from a large database.

KEY WORDS: SCHIZOPHRENIA / SUBSTANCE USE / FP-GROWTH / ASSOCIATION RULE / DATA MINING

65 pages

การค้นหากฎความสัมพันธ์แบบหลายมิติ ในผู้ป่วยโรคจิตเภทที่ใช้สารเสพติด
DISCOVERING OF MULTI-DIMENTIONAL ASSOCIATION RULE IN SCHIZOPHRENIA PATIENTS WITH SUBSTANCE USE

วิเชียร บุญญะประภา 5636643 EGIT/M

วท.ม. (การจัดการเทคโนโลยีสารสนเทศ)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์ : โษฑศ์รัตต ธรรมบุษดี, Ph.D., สุภาภรณ์ เกียรติสิน, Ph.D., อดิศร ลีลาสันติธรรม, Ph.D.

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาค้นหากฎความสัมพันธ์ ข้อมูลแบบหลายมิติ ของข้อมูลผู้ป่วยโรคจิตเภทที่ใช้สารเสพติด จากระบบฐานข้อมูลผู้ป่วยนอกแผนกจิตเวช โดยการบันทึกโรคของแพทย์ตามมาตรฐานรหัสโรค ICD-10 ประกอบด้วย 4 ปัจจัย ดังนี้ 1) ข้อมูลการใช้สารเสพติด 2) ระยะเวลาการใช้สารเสพติด 3) ข้อมูลทั่วไปของผู้ป่วย และ 4) อาการของโรคจิตเภท ใช้ข้อมูล 24 ชุด จากปัจจัยที่เป็นสาเหตุของโรคจิตเภทที่ใช้สารเสพติดจากการทำข้อมูลแบบหลายมิติ โดยวิเคราะห์ด้วยเทคนิค Frequent Patterns-Growth algorithm (FP-Growth) เป็นเทคนิคหนึ่งของการค้นหากฎความสัมพันธ์ พัฒนามาจาก Apriori algorithm เพื่อใช้ค้นหากฎความสัมพันธ์ของผู้ป่วยโรคจิตเภทที่ใช้สารเสพติด ผลการศึกษาครั้งนี้พบกฎความสัมพันธ์ระหว่างโรคจิตเภทและสารเสพติด 5 กฎ ดังนี้ 1) ชายไทยเป็นโรคจิตเภทที่ใช้สารเสพติดอายุระหว่าง 35-44 ปี 2) ผู้ป่วยโรคจิตเภทที่ใช้สารเสพติดอยู่ในเขตกรุงเทพฯ 3) ผู้ป่วยโรคจิตเภทมีการใช้บุหรี่มากที่สุด 4) พบผู้ป่วยโรคจิตเภทที่ใช้สารเสพติดอยู่ในช่วง 7-12 เดือน และ 5) ผู้ป่วยโรคจิตเภทมีพฤติกรรมเสพติดการใช้สารเสพติด กฎความสัมพันธ์ที่ถูกเลือกจากการศึกษานี้ มีระดับความเชื่อมั่นอย่างน้อย 90% ซึ่งสามารถยืนยันได้ว่าการสร้างข้อมูลแบบหลายมิติ โดยวิเคราะห์ด้วยเทคนิค FP-Growth algorithm พิสูจน์ให้เห็นว่าอาจจะเป็นเทคนิคที่เหมาะสมในการสร้างรูปแบบเพื่อใช้ในการค้นหาความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่ได้

65 หน้า

CONTENTS

CONTENTS (cont.)

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

1.1 Background and problem statement

Schizophrenia is a chronic disease that requires the continuous treatment. Schizophrenia is largely incurable disease the disease also affects the life and social functioning of patients. Schizophrenia brings a negative impact on patient, family, economy and society is enormous. The epidemic of using the amphetamine and substances around the world including Thailand is major problem in many countries. The surveys the World Health Organization [1] found that regular addictive amphetamine people were more than 27 millions. For Thailand, a methamphetamine abusing and therapy from the substance are more than 300,000 people and the dumping people are more than 1,200,000. It found that the number of addicts is increasing every year. It found that schizophrenia by and the cessations of methamphetamine were association. There is no conclusive evidence that schizophrenia like symptoms that occur in many patients is affected by amphetamine abuse and substance abuse or the actual symptoms of schizophrenia. Although the study of schizophrenia patients with a history of mental and behavioral disorders due to psychoactive substance use is widespread, however, it cannot be declared that schizophrenia is caused by using substance abuse.

Previous researches have shown that schizophrenia is associated with the risk of substance used causes schizophrenia. The duration of substance abuse resulted in higher risk of schizophrenia, and also found that the behavior of substance used, sex, age, occupation and residence of the patient. The causes of schizophrenia are significantly higher. From researches different aspect, there are other interesting knowledge from the database. Data mining technique can help to find relationships and patterns form the database, which exist on large numbers. The search for relationships and patterns by data mining techniques is a technique widely used.

Therefore, data mining techniques are used in this study to support the results from various studies.

The current data mining technique applied to medical rising steadily, and association rules technique is one of the popular technique is widely used. The association rules technique was developed for association discovery between data into large databases [2]. The association rule is applied widely for various areas such as the analysis of customer behavior. The relationship of the disease in the medical treatment of a scientific experiment, and predicting natural phenomena, etc.

Therefore, the researcher interested in the study of mental and behavioral disorders due to psychoactive substance uses associated with schizophrenia. This research idea seems reasonable to use multi-dimensional data visualization which has 4 categories: 1) psychoactive substance use, 2) duration of the substance use, 3) patient demographic data, and 4) schizophrenia status using association rules discovery with FP-Growth algorithm. This dimensionizing procedure will provide the association rules for discovering the causes of schizophrenia

1.2 Objectives

To study the process of managing data from large databases, by divided the data into multiple dimensions for analyzing multiple perspectives for discovering an association rule between schizophrenia and substance use.

To study an association between schizophrenia and substance use by using data mining to study manage parameters from a large dataset.

1.3 The hypotheses of the research

1) Can analyze for the association of addiction to substances (e.g. amphetamine), relate to the Schizophrenia in patients.

2) Can select association rule technique to help decisions support for psychiatrists and psychologists to select the techniques, and help predict the occurrence of schizophrenia.

3) Amphetamine use association with schizophrenia.

1.4 Scope of the study

1) This research gathers information about patients psychiatric and patient's information on database, from January, 2003 – November, 2014 which is the research from historical data from the history of a patient who has used substances. Data from a Siriraj Hospital in Thailand.

2) This research using data mining techniques for association. The algorithm used in the FP-Growth algorithm.

3) This research using data 4 categories: 1. psychoactive substance use, 2. duration of the substance use, 3. patient demographic data, and 4. schizophrenia status for creating multi-dimension data.

1.5 Expected results

1) To provide the information about the psychiatrist to diagnose a patient of schizophrenia.

2) To provide the information for the Psychologists, Psychiatrist and those interested use the data to determine the cause and diagnosis of schizophrenia due to substance abuse and the data were used to predict the disease.

3) The results of the study should be benefited to educate the public to realize of the harmful substance abuse and prevent in order to spread of substance in Thailand.

CHAPTER II
LITERATURE REVIEW

Nowadays, a review of the relevant literature in data mining techniques used in the analysis of the most medical data. Therefore, the researcher interested in the study used data mining techniques analysis of the data association method substance abuse in the occurrence of schizophrenia. The researcher conducted a study and research of the theoretical concept and related research, summarized the essence are as follows:

2.1 Schizophrenia

2.1.1 Characteristics and causes of schizophrenia

2.1.2 Identification of subgroups of schizophrenia

2.1.3 Clinical staging of schizophrenia

2.1.4 Criteria diagnosis of schizophrenia

2.2 Type of addiction substance

2.3 International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)

2.4 Data analysis with data mining techniques

2.4.1 Overview of data mining

2.4.2 Category of data mining tasks

2.4.3 Association rule mining technique

2.4.4 Process of analyzing data (FP-growth)

2.5 Related Researches

2.1 Schizophrenia

Schizophrenia [3] is a type of psychiatric disorder which is a medical condition. There are factors from the mind and society to affect patients with disorders of thought, mood, perception and behavior. The malfunction is unknown exactly yet. The majority of patients with onset during the adolescent, it is usually not curable. The most common symptoms have recurred periodically, which are some remains during symptoms. The exacerbate symptoms are positively syndrome, such as hallucinations and delusions. In the latter days, there is mostly negative syndrome, such as speaking a little, tardy, isolate ourselves and private world. The sex scale factor between women and men often found characteristic. Male have symptoms, are younger than females. Age of onset is found all in the 15-54 years old. Most common onset is in the adolescent or adult.

2.1.1. Characteristics and causes of schizophrenia

Clinical symptoms of schizophrenia may be divided into two major groups. Some reports may be divided into three groups; this research is into two major classes of positive symptoms and negative syndrome.

1. Positive Symptoms

Expression of the disorder of thought, perception, communication and behavior, which is divided into two sides with the symptoms.

1) Psychotic dimension such as delusions and hallucinations

2) Disorganization dimension such as disorganized behavior and disorganized speech

- Delusion common type is persecutory delusion and delusions of reference including delusions that are the main symptoms of Schneider. Impaction clinical to diagnosis is bizarre delusion, the delusions that looks weird do not understand or impossible, delusions that are the main symptoms of Schneider classified as bizarre delusion

- Hallucination, frequently is the auditory hallucination, may be a voice about patients, the sound to criticize the patient or instructed to follow. The symptoms of hallucinations are impaction to diagnosis auditory hallucination that main symptoms of Schneider

- Disorganized speech, the patient can't conclude his thoughts that will show a dialogue, must be not communicate with others effectively, such as loose associations, incoherent speech or tangentially.
- Disorganized behavior is unusually the patient dress dirty such as wear many cloths with hot weather, central public urination, some people suddenly cried out without anything stimulation

2. Negative Symptoms

A lacking of something that normal people such as in terms of feeling, the needs of the various, these symptoms include:

- Alogia : to speak a little are little speech content, take longer to answer
- Affective flattening: reduction of the expression of emotions, the look indifferent
- Avolition: lack of enthusiasm, stagnation, don't care about dressing, the patient may sit around all day doing nothing
- Asociality: expression rarely or there is an enjoyable activity in exacerbation symptoms; the poster symptoms are most common. The negative symptoms are often found in the later stages of the disease.

Biological evidence in schizophrenia supports the current concept a lot. In current camp, schizophrenia is caused by disorders of the brain and the neurodevelopment disorders caused by various factors, in each age together. There may be risk of genetic to partial will not show abnormalities, if there are no other common factors such as maternal malnutrition during pregnancy, infection in the second trinesty of pregnancy, birth injury, and mother substance abuse.

2.1.2. Subgroups of schizophrenia

According to DSM-IV [4] Schizophrenia symptoms and classified as 5 subgroups, founding in each range of illness may have altered between substance include:

- Paranoid type, the important characteristic is an obsession to delusions which is systematized delusion, disorganized or catatonic behavior or if there is no

distinctive symptoms. The most common type Patients usually begins as a small group over another. Overall inviolate personality as other subgroups.

- Disorganized type, an important characteristic is displacement or not the same direction of expressed in words or expression such as incoherence, loosening of association, symptoms are generally chaotic behavior, misunderstand, delusion that there is content not relevant. The show of the mood is straight or not appropriate apparently. Patients often how symptoms in young age. There are negative symptoms associate when illness is a long time symptoms.

- Catatonic type, a group with outstanding movement disorders includes stupor, negativism, rigidity, excitement or posturing. There are good results of treatment in this symptoms, so there symptoms are handle found in nowadays.

- Undifferentiated type, a patient with symptoms compatible with schizophrenia, but can be grouped into three clearly subgroups as any of the above.

- Residual type, a patients with relapsing symptoms had at least one clear once in a while does not evaluate positive symptoms, or may have symptoms but not outstanding. The remaining symptoms are nest of the negative symptoms these patients may be called as a group in partial remission.

2.1.3. Clinical staging of schizophrenia

A series of diseases, which is typically divided into three phases, as follows:

- Prodromal phase, most patients will gradually changing. The affected problem is responsibilities, relationships, studying, working which is getting worse. Note that the same relative inactivity went with a friend as ever keep in my room neglect of hygiene or dressing. He may become interested in the field of philosophy, psychology, religion, or to use a word or phrase of sorts. Sometimes there are behaviors that seem odd. May issue a little paranoid but it does not clearly wrong. Relatives or close ones that patients tend to not like the original.

During this uncertainty period of uncertainty is generally more difficult to detect which is approximately 1 year average before relapsing. The worsen prognosis is in long term patient and wisely patient look worse.

- Active phase, as long as the patient has symptoms. A diagnostic criterion is majority of the positive symptoms.

- Residual phase, symptoms similar to those Onset of symptoms may be flat affect or deteriorate over psychotic symptoms such as hallucinations or delusions may persist, but did not affect the patients as early found that 25 percent of schizophrenic patients with depression as well. And most of this period.

2.1.4. Criteria diagnosis of schizophrenia

According to the diagnostic criteria for schizophrenia as a person must be a qualified as follows:

- There are 2 or more of the following symptoms, which there are symptoms more than 1 month.

1. Delusions
2. Hallucination
3. Confused talking, always misunderstand and quily
4. Continually and aim less by strange behavior
5. Negative symptoms including indifference emotional, inaction or not talking.

Doctors also are researched on to say that the patient may become psychiatric disorders that are similar to these symptoms or not, this step requires experience in patient care, especially in cases where the symptoms are not clear. The experience and skills through practice and patient care, a number of other important steps are to see if the patient is a physical disease, drug or substance abuse services. Or not, as there are many physical ailments that cause symptoms similar to psychosis, substance abuse, for example, we may obviously amphetamine addicts have dementia, paranoia, fear of harm. People who eat some pills have this as well, although not as severe as in the case where it is deemed necessary to make special doctor. Such as blood levels of various substances. In the body, the brain waves or computer tomography, etc.

2.2 Type of addiction substance

1. Amphetamines

Methamphetamines [4] is a stimulant substance, discovered by Nagai in 1887, from substances in Ma-Huang is characterized Ephedrine was used in bronchodilator, is peripheral effect and report amphetamine users. In that more wrong is a more up to the present. Especially in matters relating to the neurotransmitters effects on brain and behavior, using in medical treatment and the occurrence Psychosis from amphetamines

- Central effects

Affect the release and reuptake catecholamine, and also the destruction by the enzyme monoamine oxidase in data support is the use of reserpine. That reduce the accumulation catecholamine in synaptic vesicles has no impact on amphetamine and when used alpha-methyl tyrosine is inhibited the creation. Inhibited effect of amphetamine brain stimulation.

1.) Peripheral effects

Amphetamines have a stimulating effect on alpha and beta-receptors.

(1) Cardiovascular

- Increase high blood pressure in both systolic and diastolic
- Low heart beat, when using small quantities of amphetamine
- Rarely found taehycodia, pal potation, arrhymia, but they are also found in large quantities using.

(2) Thermal regulation

Effects can also cause peripheral hyperthermia through the stimulation of the sympathoadrenal system.

(3) Gastrointestinal

- Cause constipation due to decreasing of bowel movements
- Gastric emptying time is longer to effect absorption of other substance

2.) Behavioral effects

(2.1) Locomotion

Locomotors activity was stimulated through the nucleus accumbency through dopaminergic system, but some studies have shown that it has a stimulating effect neostriatum but less than the nucleus accumbency.

(2.2) Stereotypies

A high dose of amphetamine use Stimulating affect, but less stereotyped behavior is different in each species behavior is repeated continuously, but pointless. As the experiments in mice have a bite, the smell, but not found in person.

(2.3) Aggression

The aggressive behavior is related about consumption, environment and people. Mostly are results of amphetamine intoxication and sometimes from a paranoid delusion of amphetamine psychosis.

(2.4) Anorexia

According to amphetamine that can decrease appetite. It had been used in the treatment of obesity.

- Amphetamine related dependence

According DSM-IV amphetamines could induce psychiatric disorders. Now is focus on the Amphetamine intoxication, Amphetamine withdrawal, Amphetamine delirium and Amphetamine psychotic disorder.

- Amphetamine intoxication

Amphetamine causes dopamine releasing. So the result is increasing of dopamine in the large quantities us of amphetamine cause impair judgment, impulsiveness, hyper sexuality, compulsively repeated action, hyper vigilance extreme psychomotor activation, seizure, coma.

From a study of Gershon and Angrist (1970) found that the quantity that causes psychosis is different. Most of using the amphetamine = 955 mg, but use 100 mg 1 only one home symptoms of psychosis intoxication, and there will disappear within 24 hours, and usually disappears within 48 hours.

- Amphetamine intoxication delirium

Usually in high dose or long term use and the effects of sleep deprivation. That also caused by the combination of other substances such as alcohol or those with pre-existing brained to damage.

- Amphetamine withdrawal

Amphetamines have been reduced or stopped using for a long time and too much. Typically, symptoms will occur in 2-4 days and it will be gone within 1 week. The brain showed a decrease of the neurotransmitter, catecholamine. From especial

ling nor epinephrine Watson's, Hartmann, Schildkrant, study a.d. 1972, measured from affective state, sleep patterns and 24 hr. MHPG is metabolite of nor epinephrine urine, will be reduced, and return back to normal.

2. Alcohol

Alcohol [4] is all kind of the ethyl alcohol components, such as alcohol, beer, wine, which effect to GABA receptor which depresses nerves system. There are effect the motion control, emotion and consciousness. Alcohol is good absorption in the intestine. 90 Percents of alcohol can degrade by oxidation process to carbon dioxide and water, which is excreted through the kidneys.

- Alcohol intoxication

Symptoms are not associated with a blood alcohol level only. However, there are depending on the rate of increasing in alcohol level. The faster of alcohol level is happened the more severe symptoms. If alcohol remains in the bloodstream for long time, the results will be even lower. Even for chronic alcohol drinking has high dose of alcohol level, may be no symptoms

The relationship between blood alcohol level and symptoms as shown in Table 2.1:

Table 2.1 The relationship between blood alcohol level and symptoms.

Alcohol level (mg. /dl.)	Symptoms
0-100	Jolly fun sleepy help fall asleep.
100-150	Not co-ordination of restlessness
150-250	Slurred speech, ataxia
>250	Loss of consciousness

- Alcohol withdrawal

Those who are mild alcohol addiction show symptoms during onset of approximately 6 hours after stopping alcohol. With mild symptoms such as irritability, restlessness, shaking hands but in patients with a history of nearly drinking are potentially severe symptoms include confusion, hallucinations as shown in Table 2.2.

Table 2.2 Alcohol withdrawal symptoms.

Time of stopping/ alcohol down (hr.)	**Symptoms**	**The progression of symptoms**
6-24	Irritability, sweating, shaking, high blood pressure, flushing, insomnia, rapid heart rate nystagmus hallucination illusion.	There is a 48-72 hour 5 percent in this group with a chance of severe symptoms such as delirium tremens.
7-48	Grand mal seizure (rum fits)	Peak is about 24 hours. The most characteristic is a grand mal seizure, focal seizure is found outside it.
48-72	Delirium tremens, Disorientation hallucination, insomnia, fever, confusion.	There are symptoms about 72 hr., then gradually decrease in symptoms as 5-10 days. The server symptoms one 4-5 days of stopping drinking, also medical condition such as live failure, pneumonia, GI bleeding, electrolyte

3. Cocaine

Cocaine [4] is refers to substances that cause problems to reinforce the worst of the treatment. In Thailand, cocaine use among those with higher socioeconomic status.

The mechanism of action is like amphetamine. Relatively short duration of action. Because the half-life is 30-90 minutes. To take effect faster is injected into a vein. Inhalation or nose is the most popular.

- Cocaine intoxication

Psychological symptoms include restlessness, agitation, anxiety, manic-like symptoms; physical symptoms include faster heart beat, high blood pressure, papillary dilation, stereotyped movement, confusion, delirium, and seizures.

- Cocaine withdrawal

Symptoms can occur within the first 3 days, starting with agitation, dysphoria, depression, anorexia, high cocaine craving, later be fatigue. A lot of sleeping, which is intermittently wake up, along with a good eating. This time, depression may be severe, and there is risk of suicide. In late stage, eating and sleeping will return more common. Craving is reduced, but it was anxiety, cocaine craving easily. If patient found to stimuli associated with cocaine use before.

4. Cannabis

Cannabis [4] there is unknown mechanism. Active metabolites of marijuana are tetrahydrocannabinol (THC) has a half-life up to 50 hours. Due to redistribute the fat quickly, the duration of action after the first time use, only 2-3 hours to make a jolly mood, appetite, suppression concerns, affect control of movement and the various decisions.

- Cannabis intoxication

Symptoms are impaired motor coordination, mood, anxiety, feeling delightfully slower time, impaired decision and separation while using or have used marijuana soon. Within 2 hours of using marijuana may be palpitations, dry mouth, and appetite suppression.

- Chronic cannabis syndrome (motivational syndrome)

Using high-dose, continuous cannabis causes the symptoms of apathy (lack emotion), diminished goal-directed activity also called the motivational syndrome. Patients can not deal with any problems that occur. There is separation personality. The decision is decline and deficiencies the ability to communicate.

5. Volaties

Volaties [5] means substances that can be volatile substances such as cleaning solvents, gas, gasoline, nail polish remover, and spray adhesive.

- Action

Action by penetration into the lungs affecting the brain, the perception is wrong.

- Effection

The patients are faint peace of mind and have perception change from the usual, lack of oxygen cause intoxication and hallucinations.

- Symptoms

It will cause nervous system symptoms such as loss of cardiac, excitement, delightfully, aggressive, sexual aggression, there may be died due to cardiac arrest, respiratory depression, seizures, chest pain, if the patient use it in long time, will destruct nervous system permanently and died.

2.3 International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)

Diagnostic and Statistical Manual of Mental Disorder, fourth edition (DSM-IV, published in 1994, revised in 2000 and for review in 2010), which was prepared by WHO and International Classification of Disease (ICD-10)

Table 2.3 Mental and behavioural disorders due to psychoactive substance use (F10-F19).

Code	Characteristics of the clinic
F10	Mental and behavioural disorders due to use of alcohol
F11	Mental and behavioural disorders due to use of opioids
F12	Mental and behavioural disorders due to use of cannabinoids
F13	Mental and behavioural disorders due to use of sedatives or hypnotics
F14	Mental and behavioural disorders due to use of cocaine
F15	Mental and behavioural disorders due to use of other stimulants, including caffeine (Amphetamine)
F16	Mental and behavioural disorders due to use of hallucinogens
F17	Mental and behavioural disorders due to use of tobacco (Nicotine)
F18	Mental and behavioural disorders due to use of volatile solvents
F19	Mental and behavioural disorders due to multiple drug use and use of other psychoactive substances

Table 2.4 The fourth-character subdivisions use with categories F10-F19.

Code	Characteristics of the clinic
.0	Acute intoxication
.1	Harmful use
.2	Dependence syndrome
.3	Withdrawal state
.4	Withdrawal state with delirium
.5	Psychotic disorder
.6	Amnesic syndrome
.7	Residual and late-onset psychotic disorder
.8	Other mental and behavioural disorders
.9	Unspecified mental and behavioural disorder

Table 2.5 Substance of schizophrenia (F20).

Code	Characteristics of the clinic
F20.0	Paranoid schizophrenia
F20.1	Hebephrenic schizophrenia
F20.2	Catatonic schizophrenia
F20.3	Undifferentiated schizophrenia
F20.4	Post-schizophrenic depression
F20.5	Residual schizophrenia
F20.6	Simple schizophrenia
F20.8	Other schizophrenia
F20.9	Schizophrenia, unspecified

2.4 Data Analysis with data mining techniques

2.4.1 Overview of data mining

Data mining is a technique to analyze the data which aims to find of knowledge the unknown patterns or rules of a large database. To analysis of information or knowledge to be utilized in the decision, using computer technology. Data mining [6] [7] [8] is applied use in to medical and health care. The resulted in a successful adoption. That data mining techniques can help analyze medical information.

However, theoretical, and some experts have provided a meaning of data mining Linoff, Gordon S., and Michael JA Berry states that “The exploration and analysis of large quantities of data in order to discover meaningful patterns and rules” [9].

Moreover Han, Jiawei, and Micheline Kamber also states that “Extraction of interesting (non-trivial, previously, unknown and potential useful) information from data in large databases” [10].

From both experts can be concluded that data mining is "to find something useful from a large database."

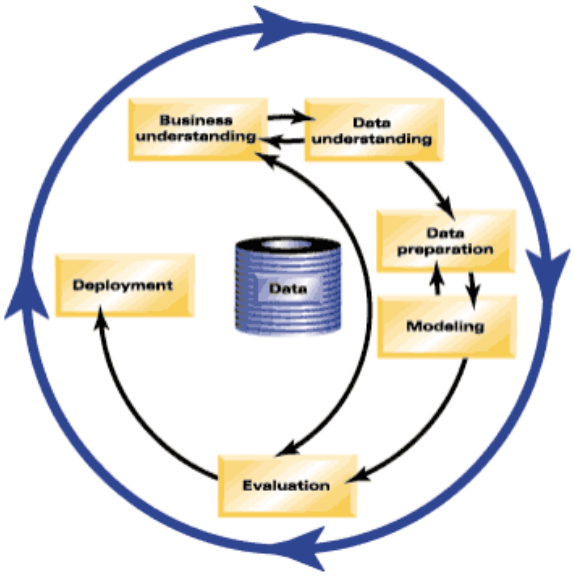


Figure.2.1 CRISP-DM Process.

The data standard, which is widely used as a blueprint. This is standard procedure for the analysis of data mining was developed in 1996 by a partnership of 3 companies as DaimlerChrysler, SPSS and NCR process called. "Cross-Industry Standard Process for Data Mining" or abbreviated as "CRISP-DM".

According processed CRISP-DM [11] it consists of 6 steps, each step is a step that is continuous, that is the next step is waiting for the result from the previous step, which is shown by the arrows between each rectangle box.

1. Business Understanding

As a first step in the process which CRISP-DM [12] which aims to understand the issues and problems that have converted to the format of the data analysis. The mining plan in action shortly.

- Consists of the sub-process as follows.

1) The goal of data mining so as to increase sales

2) Assessment of the situation in various aspects like adequacy of basic knowledge on the subject to data mining. The benefits derived from data mining worth the cost of the lost

3) Set the goal of data mining. Unlike a primary focus on solutions. Planning data mining in data collection methods and selection algorithms to be used in data mining.

2. Data Understanding

Is to identify the data source that will be used in data mining, including the required information out of the database to be taken into account in the preliminary. This step starts to collect data. Afterwards, will verify the information that has been gathered to show the accuracy of the information and determine whether to use all or part of the data is required to be used in the analysis.

- Consists of the sub-process as follows.

1) Data collection

2) Determine the properties of the data collected.

3) A brief survey of the various statistical data

4) Check the basic information on both the accuracy and integrity of information.

3. Data Preparation

The preparation of data by extracting information that is information which the recording was wrong. Information is recorded data is complex or inconsistencies, and then collect the required data from multiple databases. Intended to ensure that the quality of information that is selected is appropriate. Complete information is used to ensure that data mining can be done. Indeed, most organizations are not prepared for data mining in particular. The data is taken from various sources. Poorly stored the information is taken from the outside and brings order to the information contained within. The main problem of data quality and data integrity.

In this process it is necessary to convert the raw data into information is information that can be analyzed in the next step. By converting this information may need to be doing the same data cleaning, data conversion scale to fill the missing information. This process is a process that takes most of the CRISP-DM.

- Consists of the sub-process as follows.

1) Selection of data to be used

2) Modified form information

3) Data cleansing. Be prepared to be used in the most appropriate next steps, such as dealing with the availability of information. So that the information is accurate, complete, modify the appropriate value in the decision, to choose the data of interest. Remove all rows and columns have the same value can’t be issued because the format of the data

4. Modeling

This step is a step in the data analysis with data mining techniques, which at this stage, many of the techniques will be used to get the best response. So sometimes you have to go back to step 3 data preparation to convert some of the data to match each technique.

- Consists of the sub-process as follows.

1) Select the appropriate algorithms in data mining.

2) Customize the test results.

3) Based modeling algorithm selected.

4) Test model is derived that is accurate and reliable.

5. Evaluation

This step is the analysis of data and technical data mining, but before the results use will also have to measure the performance of results that meet the objectives that are set in the first step. There is much more reliable. This can be traced back to the previous step to change to get the desired result. To be revised to enable more efficiently before the actual implementation.

6. Deployment

The process of the CRISP-DM is not the end, just the results of the data analysis technique only mining although the results are shown that knowledge helpful. But it must be the knowledge that these practical organization or company.

2.4.2 Category of data mining tasks

1.) Association Rule Discovery

Association analysis is the discovery of what are commonly called association rules. It studies the frequency of items occurring together in transactional databases, and based on a threshold called support, identifies the frequent item sets. Another threshold, confidence, which is the conditional probability than an item appears in a transaction when another item appears, is used to pinpoint association rules. Association analysis is commonly used for market basket analysis.

2.) Classification

Classification analysis is the organization of data in given classes. Also known as supervised classification, the classification uses given class labels to order the objects in the data collection. Classification approaches normally use a training set where all objects are already associated with known class labels. The classification algorithm learns from the training set and builds a model. The model is used to classify new objects.

3.) Clustering

Similar to classification, clustering is the organization of data in classes. However, unlike classification, in clustering, class labels are unknown and it is up to the clustering algorithm to discover acceptable classes. Clustering is also called unsupervised classification, because the classification is not dictated by given class labels. There are many clustering approaches all based on the principle of maximizing

the similarity between objects in a same class (intra-class similarity) and minimizing the similarity between objects of different classes (inter-class similarity).

2.4.3. Association Rule Mining Technique

The association rule mining technique help derive valuable relations within the data points. In our case, such association will help us recognize pattern in a particular application. The association a rule composes of two item sets called an antecedent and consequent. Antecedent is the preceding event and Consequent is an event associated and followed after antecedent. In other words, an event occurring due to antecedent is followed by the consequent is depicted by the association rules for a particular class. Each rule is associated with three parameters. 1) Support is the percentage of transactions that the rule can be applied to (the percentage of transactions, in which it is correct). 2) Confidence is the number of cases in which the rule is correct relative to the number of cases in which it is applicable (and thus is equivalent to an estimate of the conditional probability of the consequent of the rule given its antecedent). 3) Lift is the ratio of the probability that antecedent and consequent occur together, to the multiple of the two individual probabilities for antecedent and consequent. There are conditions where both support and confidence is high, and still result in to invalid rule. Therefore Lift indicates the strength of a rule over random occurrence of antecedent and consequent, given their individual support. It provides information about improvement and increase in probability of consequent for a given antecedent. In the other words, Lift is given as Lift = Rule Support Support (Antecedent) ∗ Support (Consequent). 4) Conviction is not a symmetric measure. A conviction around 1 says that X and Y are independent; while conviction is infinite as conf (X→Y) is tending to 1. Note that if P(Y) is high, 1 − P(Y) is small. In that case, even if conf(X, Y) is strong, conviction (X⇒Y) may be small

2.4.4. Process of analyzing data (FP-growth)

Many algorithms for generating association rules have been presented over time, such as the Apriori and FP-Growth algorithms [13] [14]. A common strategy that these algorithms implement, in terms of performance improvement, is to decompose the problem into two subtasks:

a) Frequent item set generation, accomplished by reducing either the number of:

i. Candidate item sets based on the support measure, as in the Apriori algorithm, or

ii. Comparisons, as in the FP-Growth algorithm;

b) rule generation, which first excludes rules that have empty antecedents or consequents and then checks that, after splitting item set Y into two non-empty subsets (X and Y – X), rule X -> Y – X satisfies the confidence threshold. Y – X in this case is known as the rule consequent. Rule generation does not require any additional passes over the dataset.

The FP-Growth algorithm encodes the input data set into a compact data structure known as an FP-tree. In certain data sets, the FP-Growth algorithm outperforms the standard Apriori algorithm by several orders of magnitude, depending on the compaction factor of the FP-tree.

The FP-tree is constructed by reading the data set one transaction at a time and mapping each to a path in the tree. Paths in the FP-tree overlap when different transactions share common items. The more paths overlap, the greater the compression achieved with the FP-tree structure. If the size of the FP-tree is small enough, it will fit in main memory, from which frequent itemsets can be directly extracted. Otherwise repeated passes need to be made over the data on disk storage.

In building the FP-tree, one pass is made over the data to determine support count for each item, discard infrequent items, and sort the frequent ones in order of decreasing support counts. The data set is then scanned once more to read each transaction and add its corresponding path to the initial tree, which consists of simply a root node. For each transaction read, a new set of nodes is created, as long as the paths do not share a common prefix. When paths share a common prefix (same initial item), they overlap in the tree, and the support count for the shared node is incremented by one. Once every transaction has been read and mapped on a path, the resulting FP-tree is ready. The size of the resulting tree depends on the ordering of the items.

The frequent itemset generation in the FP-Growth algorithm is done in a bottom-up fashion, starting with a particular ending item. Only the paths containing

that node are examined, after ensuring that it is a frequent itemset itself. The support counts along the prefix paths are updated to count only transactions that include the node in question, as well as to truncate the paths by removing that node. Then the algorithm tries to find frequent itemsets ending in that node paired with each of the other nodes that immediately precede it in the FP-tree. This is done in a recursive fashion.

In order to avoid an unmanageable amount of rules created, it is important to clearly set criteria for evaluating the quality of association patterns. This can be done:

a) Objectively, through measures that use statistics, such as support, confidence, lift (or interested factor), and correlation, to determine the interestingness of a pattern;

b) subjectively, by using domain expertise to determine whether the information or knowledge revealed about the data is interesting.

2.5 Related Research

Schizophrenia [15] is one type of psychiatric disorders required medical treatment. An onset of schizophrenia is usually in adolescent which etiology is not clear. Symptoms of schizophrenia could be classified into three main categories: (1) positive symptoms, (2) negative symptoms, and (3) cognitive symptoms. The positive symptoms are included hallucinations, delusions, thought disorders, and movement disorders. The negative symptoms are included problem with motivation, social withdrawal, and diminished affective responsiveness, speech and movement. The cognitive symptoms are poor executive function, impair focus and attention problem, and impair working memory. Furthermore, schizophrenia has high recurrent rate and several psychological co-morbidities. Schizophrenia can be caused by several factors. Therefore, it is necessary to identify the actual cause of schizophrenia in each patient in order to provide the proper treatment for each patient.

An advanced information technology, such as data mining techniques, provided the opportunity to analyze and to predict the cause of this psychiatric disorder [16]. Performed the population-based cohort study aimed to investigate the

relationships of schizophrenia with the psychoactive substance abuse. The risk of schizophrenia in the methamphetamine abuse group was significantly higher than the appendicitis group (hazard ratio = 9.37).

The data mining technique, i.e. an applied Apriori algorithm, could be used to uncover the level of illness from the large database of depressed people [17]. There are several ways to perform the association rule discovery, e.g. a FT-Growth algorithm, an Apriori algorithm, etc. Comparatively, the FP-growth algorithm outperforms the Appriori algorithm for mining a huge number of finding rule [18].

The FP-growth Algorithm [19] is efficient and scalable for identifying the complete set of frequent patterns. Two data mining techniques, i.e. support vector machines and decision tree, were used to predict schizophrenia treatments. The decision tree technique demonstrated the better accuracy (approximately 90%) comparable to the support vector machine technique.

The predictive performance of the FP-Growth algorithm and the Apriori algorithm has been shown to be congruent a study of acute methamphetamine-induced psychosis and acute negative-methamphetamine schizophrenia patients did not find the statistically significant relationship between the blood methamphetamine concentrations and the severity of psychotic symptoms [20] [21].

However, the onset of schizophrenia might require the duration of psychoactive substance use.

CHAPTER III
RESEARCH METHODOLOGY

This study will test the applied data mining techniques by using association rule technique with FP-Growth algorithms to the non-identifiable schizophrenic patient data. The demographic data, the history of substance use, duration of usage, and behavioral disorder were extracted from the database systems and generation of the multi-dimensions categories of data for analysis.

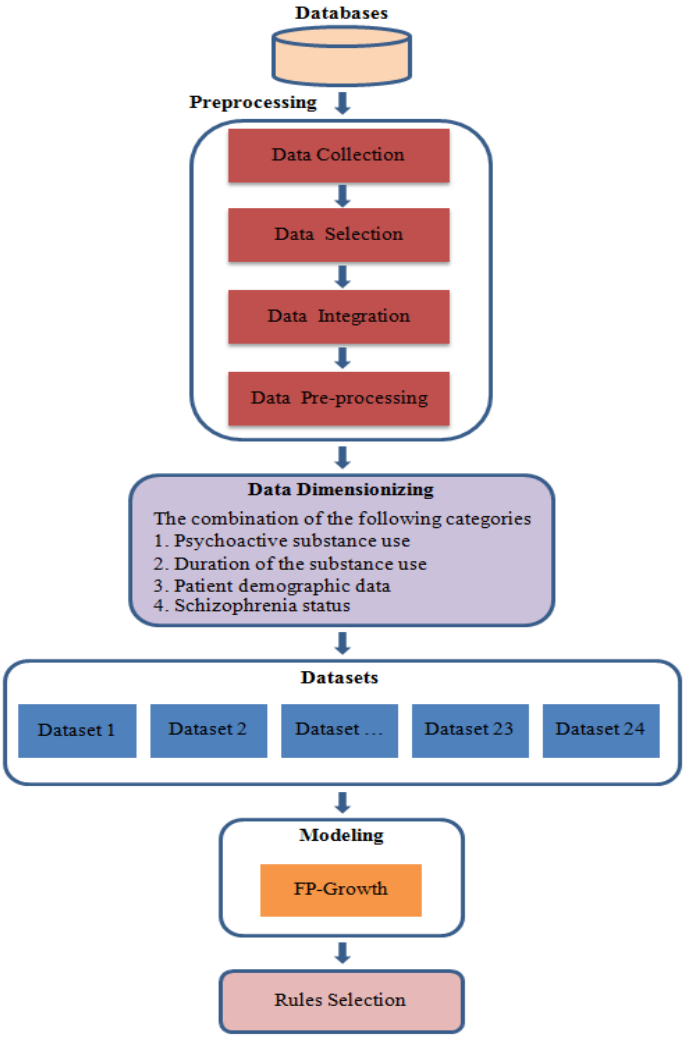


Figure III.1 Overview of the study.

In this study, FP-Growth algorithm was performed in order to determine the association rule of schizophrenia and the psychoactive substance abuse. This study processed the data analysis by (1) preprocessing, (2) data dimensionizing, (3) modeling, and (4) model evaluation, so were shown in the Figure 3.1.

3.1 Preprocessing

(1) Data collection

The dataset was collected from the psychiatric outpatient department clinic, including the record of schizophrenia (F20) and other mental and behavioral disorders due to the psychoactive substance usage (F10-19). The diseases were clinically diagnosed according to the diagnostic criteria of DSM-IV, and were classified into the database according to the ICD-10 standard codes. Be extracted from the database systems with the following criteria.

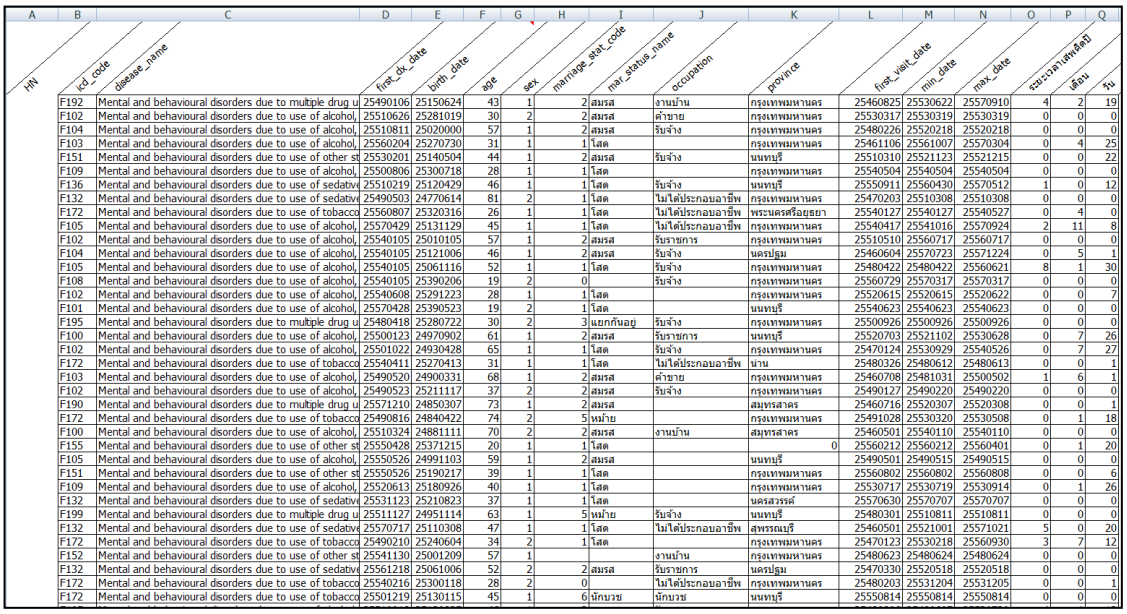
Inclusion criteria

1. The data is recorded by medical record.

2. The medical record of the schizophrenic patients and the patients who have substance related disorder who are diagnosed by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria and are recorded into the database by the International Classification of Diseases, Tenth Edition (ICD-10).

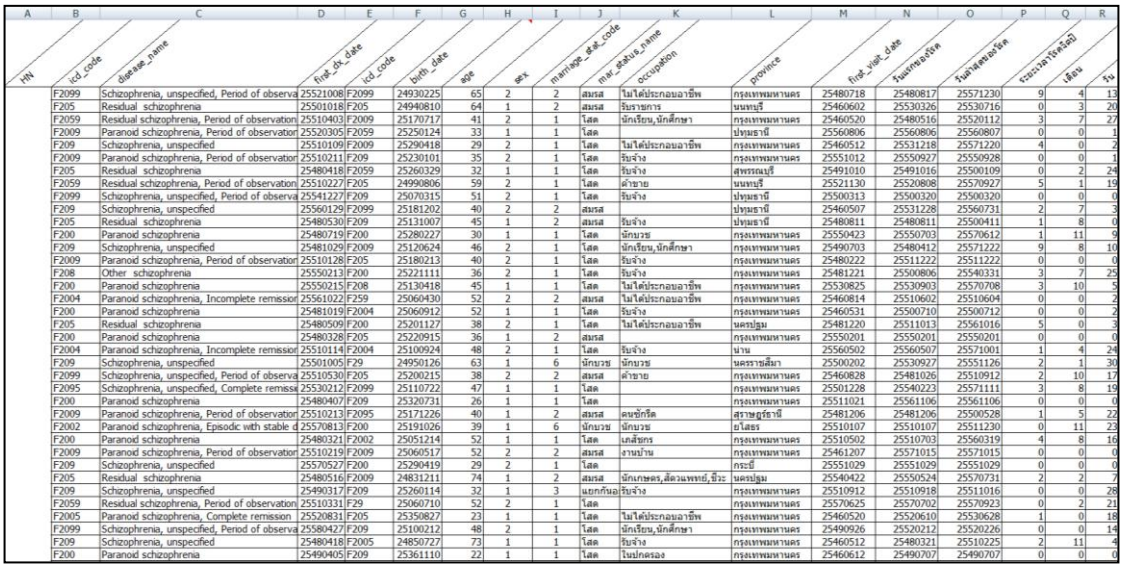
Exclusion criteria

None



HN	icd_code	disease_name	first_dx_date	birth_date	age	sex	marriage_stat_code	mar_status_name	occupation	province	first_visit_date	min_date	max_date	ระยะเวลาสังเกตสั้นสุด	เสียน	รัน
	F192	Mental and behavioural disorders due to multiple drug u	25490106	25150624	43	1	2	สมรส	งานบ้าน	กรุงเทพมหานคร	25460825	25530622	25570910	4	2	19
	F102	Mental and behavioural disorders due to use of alcohol,	25510626	25281019	30	2	2	สมรส	ค้าขาย	กรุงเทพมหานคร	25530317	25530319	25530319	0	0	0
	F104	Mental and behavioural disorders due to use of alcohol,	25510811	25020000	57	1	2	สมรส	รับจ้าง	กรุงเทพมหานคร	25480226	25520218	25520218	0	0	0
	F103	Mental and behavioural disorders due to use of alcohol,	25560204	25270730	31	1	1	โสด		กรุงเทพมหานคร	25461106	25561007	25570304	0	4	25
	F151	Mental and behavioural disorders due to use of other st	25530201	25140504	44	1	2	สมรส	รับจ้าง	นนทบุรี	25510310	25521123	25521215	0	0	22
	F109	Mental and behavioural disorders due to use of alcohol,	25500806	25300718	28	1	1	โสด		กรุงเทพมหานคร	25540504	25540504	25540504	0	0	0
	F136	Mental and behavioural disorders due to use of sedative	25510219	25120429	46	1	1	โสด	รับจ้าง	นนทบุรี	25550911	25560430	25570512	1	0	12
	F132	Mental and behavioural disorders due to use of sedative	25490503	24770614	81	2	1	โสด	ไม่ได้ประกอบอาชีพ	กรุงเทพมหานคร	25470203	25510308	25510308	0	0	0
	F172	Mental and behavioural disorders due to use of tobacco	25560807	25320316	26	1	1	โสด	ไม่ได้ประกอบอาชีพ	พระนครศรีอยุธยา	25540127	25540127	25540527	0	4	0
	F105	Mental and behavioural disorders due to use of alcohol,	25570429	25131129	45	1	1	โสด	ไม่ได้ประกอบอาชีพ	กรุงเทพมหานคร	25540417	25541016	25570924	2	11	8
	F102	Mental and behavioural disorders due to use of alcohol,	25540105	25010105	57	1	2	สมรส	รับราชการ	กรุงเทพมหานคร	25510510	25560717	25560717	0	0	0
	F104	Mental and behavioural disorders due to use of alcohol,	25540105	25121006	46	1	2	สมรส	รับจ้าง	นครปฐม	25460604	25570723	25571224	0	5	1
	F105	Mental and behavioural disorders due to use of alcohol,	25540105	25061116	52	1	1	โสด	รับจ้าง	กรุงเทพมหานคร	25480422	25480422	25560621	8	1	30
	F108	Mental and behavioural disorders due to use of alcohol,	25540105	25390206	19	2	0		รับจ้าง	กรุงเทพมหานคร	25560729	25570317	25570317	0	0	0
	F102	Mental and behavioural disorders due to use of alcohol,	25540608	25291223	28	1	1	โสด		กรุงเทพมหานคร	25520615	25520615	25520622	0	0	7
	F101	Mental and behavioural disorders due to use of alcohol,	25570428	25390523	19	2	1	โสด		นนทบุรี	25540623	25540623	25540623	0	0	0
	F195	Mental and behavioural disorders due to multiple drug u	25480418	25280722	30	2	3	แยกกันอยู่	รับจ้าง	กรุงเทพมหานคร	25500926	25500926	25500926	0	0	0
	F100	Mental and behavioural disorders due to use of alcohol,	25500123	24970902	61	1	2	สมรส	รับราชการ	นนทบุรี	25520703	25521102	25530628	0	7	26
	F102	Mental and behavioural disorders due to use of alcohol,	25501022	24930428	65	1	1	โสด	รับจ้าง	กรุงเทพมหานคร	25470124	25530929	25540526	0	7	27
	F172	Mental and behavioural disorders due to use of tobacco	25540411	25270413	31	1	1	โสด	ไม่ได้ประกอบอาชีพ	น่าน	25480326	25480612	25480613	0	0	1
	F103	Mental and behavioural disorders due to use of alcohol,	25490520	24900331	68	1	2	สมรส	ค้าขาย	กรุงเทพมหานคร	25460708	25481031	25500502	1	6	1
	F102	Mental and behavioural disorders due to use of alcohol,	25490523	25211117	37	2	2	สมรส	รับจ้าง	กรุงเทพมหานคร	25490127	25490220	25490220	0	0	0
	F190	Mental and behavioural disorders due to multiple drug u	25571210	24850307	73	1	2	สมรส		สมุทรสาคร	25460716	25520307	25520308	0	0	1
	F172	Mental and behavioural disorders due to use of tobacco	25490816	24840422	74	2	5	หม้าย		กรุงเทพมหานคร	25491028	25530320	25530508	0	1	18
	F100	Mental and behavioural disorders due to use of alcohol,	25510324	24881111	70	2	2	สมรส	งานบ้าน	สมุทรสาคร	25460501	25540110	25540110	0	0	0
	F155	Mental and behavioural disorders due to use of other st	25550428	25371215	20	1	1	โสด		0	25560212	25560212	25560401	0	1	20
	F105	Mental and behavioural disorders due to use of alcohol,	25550526	24991103	59	1	2	สมรส		นนทบุรี	25490501	25490515	25490515	0	0	0
	F151	Mental and behavioural disorders due to use of other st	25550526	25190217	39	1	1	โสด		กรุงเทพมหานคร	25560802	25560802	25560808	0	0	6
	F109	Mental and behavioural disorders due to use of alcohol,	25520613	25180926	40	1	1	โสด		กรุงเทพมหานคร	25530717	25530719	25530914	0	1	26
	F132	Mental and behavioural disorders due to use of sedative	25531123	25210823	37	1	1	โสด		นครสวรรค์	25570630	25570707	25570707	0	0	0
	F199	Mental and behavioural disorders due to multiple drug u	25511127	24951114	63	1	5	หม้าย	รับจ้าง	นนทบุรี	25480301	25510811	25510811	0	0	0
	F132	Mental and behavioural disorders due to use of sedative	25570717	25110308	47	1	1	โสด	ไม่ได้ประกอบอาชีพ	สุพรรณบุรี	25460501	25521001	25571021	5	0	20
	F172	Mental and behavioural disorders due to use of tobacco	25490210	25240604	34	2	1	โสด		กรุงเทพมหานคร	25470123	25530218	25560930	3	7	12
	F152	Mental and behavioural disorders due to use of other st	25541130	25001209	57	1			งานบ้าน	กรุงเทพมหานคร	25480623	25480624	25480624	0	0	0
	F132	Mental and behavioural disorders due to use of sedative	25561218	25061006	52	2	2	สมรส	รับราชการ	นครปฐม	25470330	25520518	25520518	0	0	0
	F172	Mental and behavioural disorders due to use of tobacco	25540216	25300118	28	2	0		ไม่ได้ประกอบอาชีพ	กรุงเทพมหานคร	25480203	25531204	25531205	0	0	1
	F172	Mental and behavioural disorders due to use of tobacco	25501219	25130115	45	1	6	นักบวช	นักบวช	นนทบุรี	25550814	25550814	25550814	0	0	0

Figure 3.2 Example of data of patient with mental and behavioral disorders due to psychoactive substance use (F10-F19).



HN	icd_code	disease_name	first_dx_date	icd_code	birth_date	age	sex	marriage_stat_code	mar_status_name	occupation	province	first_visit_date	วันแรกของปีงบ	วันสุดท้ายของปีงบ	ระยะเวลาสังเกตสั้นสุด	เสียน	รัน
	F2099	Schizophrenia, unspecified, Period of observa	25521008	F2099	24930225	65	2	2	สมรส	ไม่ได้ประกอบอาชีพ	กรุงเทพมหานคร	25480718	25480817	25571230	9	4	13
	F205	Residual schizophrenia	25501018	F205	24940810	64	1	2	สมรส	รับราชการ	นนทบุรี	25460602	25530326	25530716	0	3	20
	F2059	Residual schizophrenia, Period of observation	25510403	F2059	25170717	41	2	1	โสด	นักเรียน,นักศึกษา	กรุงเทพมหานคร	25460520	25480516	25520112	3	7	27
	F2009	Paranoid schizophrenia, Period of observation	25520305	F2059	25250124	33	1	1	โสด		ปทุมธานี	25560806	25560806	25560807	0	0	1
	F209	Schizophrenia, unspecified	25510109	F2009	25290418	29	2	1	โสด	ไม่ได้ประกอบอาชีพ	กรุงเทพมหานคร	25460512	25531218	25571220	4	0	2
	F2009	Paranoid schizophrenia, Period of observation	25510211	F209	25230101	35	2	1	โสด	รับจ้าง	กรุงเทพมหานคร	25551012	25550927	25550928	0	0	1
	F205	Residual schizophrenia	25480418	F2059	25260329	32	1	1	โสด	รับจ้าง	สุพรรณบุรี	25491010	25491016	25500109	0	2	24
	F2059	Residual schizophrenia, Period of observation	25510227	F205	24990806	59	2	1	โสด	ค้าขาย	นนทบุรี	25521130	25520808	25570927	5	1	19
	F2099	Schizophrenia, unspecified, Period of observa	25541227	F209	25070315	51	2	1	โสด	รับจ้าง	ปทุมธานี	25500313	25500320	25500320	0	0	0
	F209	Schizophrenia, unspecified	25560129	F2099	25181202	40	2	2	สมรส		ปทุมธานี	25460507	25531228	25560731	2	7	3
	F205	Residual schizophrenia	25480530	F209	25131007	45	1	2	สมรส	รับจ้าง	ปทุมธานี	25480811	25480811	25500411	1	8	0
	F200	Paranoid schizophrenia	25480719	F200	25280227	30	1	1	โสด	นักบวช	กรุงเทพมหานคร	25550423	25550703	25570612	1	11	9
	F209	Schizophrenia, unspecified	25481029	F2009	25120624	46	2	1	โสด	นักเรียน,นักศึกษา	กรุงเทพมหานคร	25490703	25480412	25571222	9	8	10
	F2009	Paranoid schizophrenia, Period of observation	25510128	F205	25180213	40	2	1	โสด	รับจ้าง	กรุงเทพมหานคร	25480222	25511222	25511222	0	0	0
	F208	Other schizophrenia	25550213	F200	25221111	36	2	1	โสด	รับจ้าง	กรุงเทพมหานคร	25481221	25500806	25540331	3	7	25
	F200	Paranoid schizophrenia	25550215	F208	25130418	45	1	1	โสด	ไม่ได้ประกอบอาชีพ	กรุงเทพมหานคร	25530825	25530903	25570708	3	10	5
	F2004	Paranoid schizophrenia, Incomplete remissior	25561022	F259	25060430	52	2	2	สมรส	ไม่ได้ประกอบอาชีพ	กรุงเทพมหานคร	25460814	25510602	25510604	0	0	2
	F200	Paranoid schizophrenia	25481019	F2004	25060912	52	1	1	โสด	รับจ้าง	กรุงเทพมหานคร	25460531	25500710	25500712	0	0	2
	F205	Residual schizophrenia	25480509	F200	25201127	38	2	1	โสด	ไม่ได้ประกอบอาชีพ	นครปฐม	25481220	25511013	25561016	5	0	3
	F200	Paranoid schizophrenia	25480328	F205	25220915	36	1	2	สมรส		กรุงเทพมหานคร	25550201	25550201	25550201	0	0	0
	F2004	Paranoid schizophrenia, Incomplete remissior	25510114	F2004	25100924	48	2	1	โสด	รับจ้าง	น่าน	25560502	25560507	25571001	1	4	24
	F209	Schizophrenia, unspecified	25501005	F29	24950126	63	1	6	นักบวช	นักบวช	นครราชสีมา	25500202	25530927	25551126	2	1	30
	F2099	Schizophrenia, unspecified, Period of observa	25510530	F205	25200215	38	2	2	สมรส	ค้าขาย	กรุงเทพมหานคร	25460828	25481026	25510912	2	10	17
	F2095	Schizophrenia, unspecified, Complete remissi	25530212	F2099	25110722	47	1	1	โสด		กรุงเทพมหานคร	25501228	25540223	25571111	3	8	19
	F200	Paranoid schizophrenia	25480407	F209	25320731	26	1	1	โสด		กรุงเทพมหานคร	25511021	25561106	25561106	0	0	0
	F2009	Paranoid schizophrenia, Period of observation	25510213	F2095	25171226	40	1	2	สมรส	คนขับรถ	สุราษฎร์ธานี	25481206	25481206	25500528	1	5	22
	F2002	Paranoid schizophrenia, Episodic with stable d	25570813	F200	25191026	39	1	6	นักบวช	นักบวช	ยโสธร	25510107	25510107	25511230	0	11	23
	F200	Paranoid schizophrenia	25480321	F2002	25051214	52	1	1	โสด	เกสัชกร	กรุงเทพมหานคร	25510502	25510703	25560319	4	8	16
	F2009	Paranoid schizophrenia, Period of observation	25510219	F2009	25060517	52	2	2	สมรส	งานบ้าน	กรุงเทพมหานคร	25461207	25571015	25571015	0	0	0
	F209	Schizophrenia, unspecified	25570527	F200	25290419	29	2	1	โสด		กระบี่	25551029	25551029	25551029	0	0	0
	F205	Residual schizophrenia	25480516	F2009	24831211	74	1	2	สมรส	นักเกษตร,สัตวแพทย์,ปีระ	นครปฐม	25540422	25550524	25570731	2	2	7
	F209	Schizophrenia, unspecified	25490317	F209	25260114	32	1	3	แยกกันอยู่	รับจ้าง	กรุงเทพมหานคร	25510912	25510918	25511016	0	0	28
	F2059	Residual schizophrenia, Period of observation	25510331	F29	25060710	52	2	1	โสด		กรุงเทพมหานคร	25570625	25570702	25570923	0	2	21
	F2005	Paranoid schizophrenia, Complete remission	25520831	F205	25350827	23	1	1	โสด	ไม่ได้ประกอบอาชีพ	กรุงเทพมหานคร	25460520	25520610	25530628	1	0	18
	F2099	Schizophrenia, unspecified, Period of observa	25580427	F209	25100212	48	2	1	โสด	นักเรียน,นักศึกษา	กรุงเทพมหานคร	25490926	25520212	25520226	0	0	14
	F209	Schizophrenia, unspecified	25480418	F2005	24850727	73	1	1	โสด	รับจ้าง	กรุงเทพมหานคร	25460512	25480321	25510225	2	11	4
	F200	Paranoid schizophrenia	25490405	F209	25361110	22	1	1	โสด	ไนปกครอง	กรุงเทพมหานคร	25460612	25490707	25490707	0	0	0

Figure 3.3 Example data of patient with substance of schizophrenia (F20).

(2) Data Selection

The schizophrenia and other mental and behavioral disorder datasets consisted of the following attributes: ICD-10 code, city, sex (male or female), age at diagnosis, marital status (single, married, or divorced), and substance use (yes or no). The dataset was organized by combining the schizophrenia and other mental and

behavioral disorder into the same format in order to select the patients who have them concurrently. The missing data such as the lack of any attribute data would be excluded from the dataset.

(3) Data Integration

The data was joined by HN number as the primary key between substances used (F10-F19) and schizophrenia patients (F20) for processing the dataset.

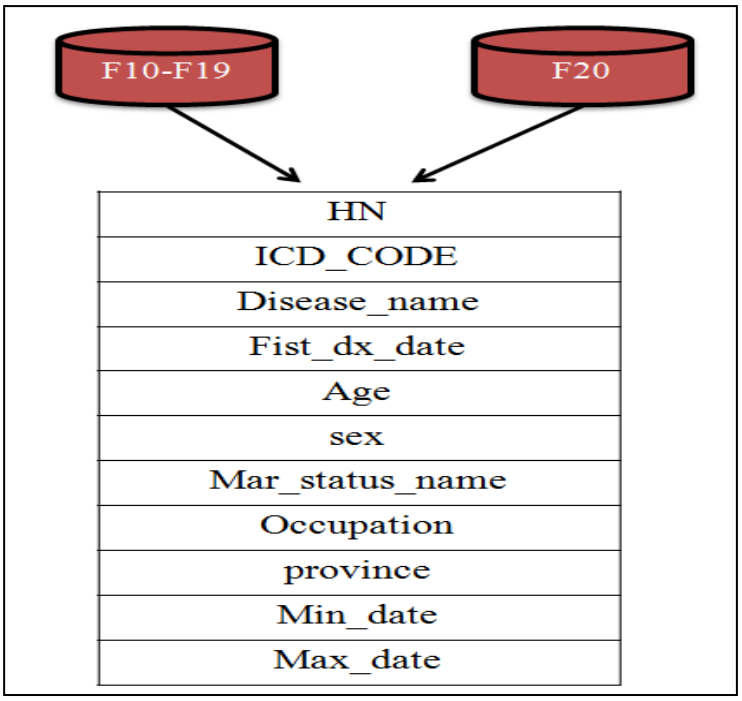


Figure 3.4 Data Integration.

(4) Data Pre-processing

The schizophrenia and other mental and behavioral disorder datasets consisted of the following attributes: ICD-10 code, city, sex (male or female), age at diagnosis, marital status (single, married, or divorced), and substance abused (yes or no). The dataset was organized by combining the schizophrenia and other mental and behavioral disorder into the same format in order to select the patients who have them concurrently. The missing data such as the lack of any attribute data would be excluded from the dataset.

The initial step of the preliminary data processing was to create a few additional attributes as computed fields. Some of these calculated fields were for the

purpose of filling voids in the dataset, by adding necessary additional attributes, such as calculating duration of substance use from years. To categorize data or encode attribute values as numbers, the only type of data the FP-Growth algorithm application used can process. In order for these numbers to be told apart in the files that are output by the FP-Growth algorithm, such data transformation is a typical step in the data mining process.

3.2 Data dimensionizing

In this step, we classify the data into 24 datasets based on the combination of the following categories, including psychoactive substance abuse, duration of the substance abuse, patient demographic data, and schizophrenia status. Also subcategorized the psychoactive substance uses patients into 2 attributes. One was the type of substance the patients used; another was the symptoms, the behavioral disorder or both mental and behavioral disorders which were the consequence from substances use. Then we choose 24 datasets, relationship with schizophrenia. First 12 datasets are composed of 2 categories. Second 12 datasets are composed of 3 categories. There are describing of multi-dimensional data in Figure 3.5 and selected 24 datasets were shown in Table 3.1.

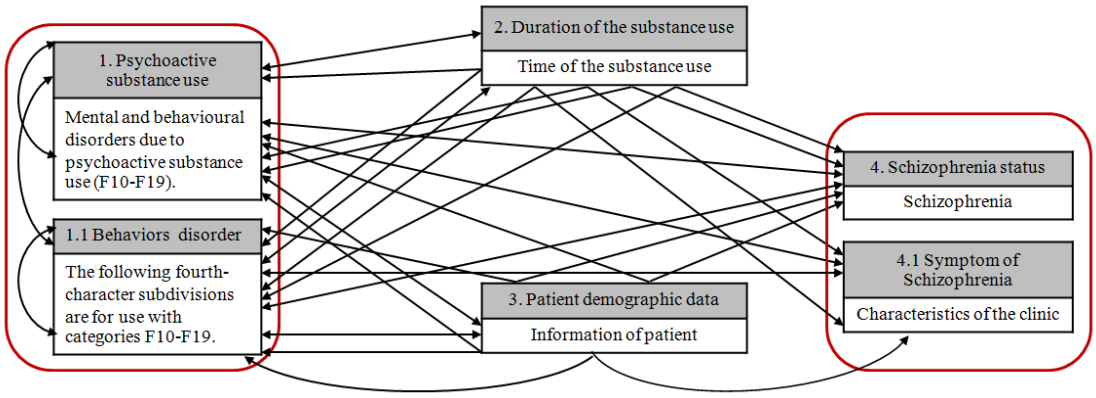


Figure 3.5 Data Multi-Dimensional Cross.

Table 3.1 Dataset descriptions.

Dataset No.	Description	Dataset No.	Description
1	Substance between Substance	13	Substance between Symptoms
2	Duration between Substance	14	Substance and duration between Symptoms
3	Social data between Substance	15	Substance and social data between Symptoms
4	Behavioural disorder between Substance	16	Behavioural disorder between Schizophrenia
5	Behavioural disorder and duration between Substance	17	Behavioural disorder and duration between Schizophrenia
6	Behavioural disorder and social data between Substance	18	Behavioural disorder and social data between Schizophrenia
7	Behavioural disorder between Behavioural disorder	19	Behavioural disorder between Symptoms
8	Duration between Behavioural disorder	20	Behavioural disorder and duration between Symptoms
9	Social data between Behavioural disorder	21	Behavioural disorder and social data between Symptoms
10	Substance between Schizophrenia	22	Mental and behavioural disorder between Schizophrenia
11	Substance and duration between Schizophrenia	23	Mental and behavioural disorder and duration between Schizophrenia
12	Substance and social data between Schizophrenia	24	Mental and behavioural disorder and social data between Schizophrenia

3.3 Modeling and Evaluation Process

(1.) Modeling

In order to demonstrate the association between schizophrenia and psychoactive substance usage, this study designed to perform the association rule discovery [22] [23]. After got the association between any attributes, the psychiatrist opinions for those associations still needed to confirm the reasonable of the discovery. To demonstrate the magnitude of the association, applied the FP-Growth algorithm [24], one of the association rule discovery methods to our dataset. Initially, we started with encoding the input dataset into a compact data structure called a frequent pattern tree (FP-tree).

(2.) Evaluation Process

The related parameters of association rule discovery measured by Support, Confidence, Lift and Conviction [25]. Support is the proportion of transactions containing the set X. Confidence is the proportion of transaction containing X which also contains Y. Lift is the ratio of the observed support of that expected if X and Y were independent. A conviction can be interpreted as the ratio of the expected frequency that X occurs without Y. Each parameter is shown in the following as shown as calculated from the equation (3.1) - (3.4).

Support is the proportion of transactions containing the set X:

$$\text{supp}\,(X) = \frac{\text{number of transactions containing\_X}}{\text{number of transactions}} \quad (3.1)$$

Confidence is the proportion of transaction containing X which also contains Y:

$$\text{conf}(X \Longrightarrow Y) = \frac{\text{supp}(XUY)}{\text{supp}(X)} \quad (3.2)$$

Lift is the ratio of the observed support to that expected if X and Y were independent:

$$\text{lift}(X \Longrightarrow Y) = \frac{\text{supp}(XUY)}{\text{supp}(X) * \text{supp}(Y)} \quad (3.3)$$

Conviction can be interpreted as the ratio of the expected frequency that X occurs without Y:

$$\text{conv}(X \Longrightarrow Y) = \frac{1-\text{supp}(Y)}{1-\text{conf}(X \Rightarrow Y)} \quad (3.4)$$

3.4 Rules selection

Choosing best association rule of each dataset was from this study to confirm by the psychiatrist. This study was used to diagnosis schizophrenia in the future.

CHAPTER IV
RESULTS AND DISCUSSION

This thesis, studies to create multi-dimensional data, which the data comprised of 4 categories, such as, psychoactive substance use, duration of substance use, patient demographic data, and schizophrenia status, from a sample of 562 specimens from the outpatient psychiatric database using data mining techniques by association rules discovery with FP-Growth algorithm to discover the association rules between schizophrenia and substance abuse.

4.1 Association rule results

The association rules of the highest confidence level of each the dataset are shown in Table 4.1, together with a lift, and conviction. Confidence, lift, and conviction are the values of probability, a measure of rule, and belief of rule respectively. The result of the association rules, including the rules of their confidence, lift, and conviction calculations are given in Appendix A.

Table 4.1 Summary of the average precision results of all experiment datasets.

Dataset No.	Association Rules	Confidence	Lift	Conviction
1	Substances, Volatile → Amphetamine	0.71	6.67	3.13
2	Substances, Duration > 90 → Alcohol	0.86	1.40	2.72
3	Unemployed, Diagnosis age range, Northeast region → Alcohol	0.97	1.58	11.66
4	Dependence syndrome, Withdrawal state, Withdrawal state with delirium → Alcohol	0.99	1.62	33.03

Table 4.1 Summary of the average precision results of all experiment datasets. (cont.)

Dataset No.	Association Rules	Confidence	Lift	Conviction
5	Duration 0-6, Dependence syndrome, Withdrawal state with delirium → Alcohol	0.99	1.61	30.31
6	Male, Dependence syndrome, Bangkok, Withdrawal state with delirium → Alcohol	0.99	1.62	47.79
7	Withdrawal state, Other mental and behavioral disorders → Dependence syndrome	0.98	1.39	13.28
8	Withdrawal state, Duration 49-54 → Dependence syndrome	0.95	1.35	6.20
9	Male, Hireling, Withdrawal state, Psychotic disorder → Dependence syndrome	0.97	1.38	10.33
10	Nicotine, Amphetamine, Substances → Schizophrenia	0.89	1.32	2.93
11	Amphetamine, Duration 7-12 → Schizophrenia	0.94	1.39	5.21
12	Male, Bangkok, Single, Nicotine, Diagnosis age range 35-44 → Schizophrenia	0.96	1.42	7.82
13	Nicotine, Alcohol, Substances → Paranoid schizophrenia	0.83	1.83	3.27
14	Nicotine, Duration 43-48 → Paranoid schizophrenia	0.90	1.98	5.45
15	Male, Amphetamine, Diagnosis age range 45-54 → Schizophrenia unspecified	0.89	2.14	5.26

Table 4.1 Summary of the average precision results of all experiment datasets. (cont.)

Dataset No.	Association Rules	Confidence	Lift	Conviction
16	Dependence syndrome, Unspecified mental and behavioral disorder → Schizophrenia	0.90	1.33	3.26
17	Psychotic disorder, Duration 19-24 → Schizophrenia	0.92	1.37	4.23
18	Male, Bangkok, Other mental and behavioral disorders → Schizophrenia	0.80	1.19	1.63
19	Dependence syndrome, Harmful use, Unspecified mental and behavioral disorder → Schizophrenia-unspecified	0.80	1.82	2.81
20	Dependence syndrome, Harmful use, Duration 7-12 → Paranoid schizophrenia	0.88	1.92	4.36
21	Male, Dependence syndrome, Single, Other mental and behavioral disorders → Schizophrenia unspecified	0.80	1.82	2.81
22	Nicotine dependence syndrome, Amphetamine dependence syndrome → Schizophrenia	0.95	1.41	6.84
23	Nicotine dependence syndrome, Duration 25-30 → Schizophrenia	0.92	1.37	4.23
24	Male, Bangkok, Single, Nicotine dependence syndrome, Diagnosis age range 35-44 → Schizophrenia	0.96	1.42	7.82

Table 4.2 Results rule selection of experimental datasets.

Dataset No.	Association Rules	Confidence	Lift	Conviction
11	Amphetamine, Duration 7-12 → Schizophrenia	0.94	1.39	5.21
12	Male, Bangkok, Single, Nicotine, Diagnosis age range 35-44 → Schizophrenia	0.96	1.42	7.82
22	Nicotine dependence syndrome, Amphetamine dependence syndrome → Schizophrenia	0.95	1.41	6.84
24	Male, Bangkok, Single, Nicotine dependence syndrome, Diagnosis age range 35-44 → Schizophrenia	0.96	1.42	7.82

The results of the 24 datasets (Table 4.1), the dataset with the consequence of "schizophrenia" in the association rules were selected to represent the results of the study (Table 4.2). The dataset of No.11, 94 percent of the patients diagnosed with amphetamine use in the duration of 7 to 12 months was likely to have schizophrenia with the measure of rule 1.39, and with the strong belief of rule 5.21. The dataset of No.12, 96 percent of the single male patients in Bangkok diagnosed with nicotine use in diagnosis at the age of 35 to 44 years old was likely to have schizophrenia with the measure of rule 1.42, and with the strong belief of rule 7.82. The dataset of No.22, 95 percent of the patients diagnosed with nicotine dependence, and amphetamine dependence were likely to have schizophrenia with the measure of rule 1.41, and with the strong belief of rule 6.84. The dataset of No.24, 96 percent of the single male patients in Bangkok diagnosed with nicotine dependence in the age of 35 to 44 years old was likely to have schizophrenia with the measure of rule 1.42, and with the strong belief of rule 7.82. The results together, the single male in Bangkok diagnosed with nicotine and/or amphetamine use and/or dependence diagnosis age range of 35 to 44 years old with the use of a period of 7 to 12 months was at associated with having schizophrenia.

4.2 Discussion

This study demonstrated the accuracy of the FP- growth algorithm as an analytical tool for the association rule discovery of the dataset of the multi-dimensional data.

The resulted of association rules discovery showed the acceptable confidence level with the value of at least 0.70 for overall association rules and of at least 0.90 for the selected association rules. This result was also corresponding to the previous finding [20], which approved the validity of the FP - growth algorithm in comparison with Apriori algorithm.

The interesting thing in this study to mention is introductory. Study the data which comprised of 4 categories: psychoactive substance use, duration of substance use, patient demographic data, and schizophrenia status. To create the multi-dimensional data between schizophrenia and substance use, which the dataset this study used data 24 datasets, for design to discover the correlation between various data. Besides, was the association rules discovery between schizophrenia and substance use. The highlight of this study is used the association rule discovery technique with FP-Growth algorithm, which is a technique two-way analysis (A-->B, B-->A).

The results of this study are consistent the articles in schizophrenia have found the cause of schizophrenia is associated with the nicotine and amphetamine. Which the substance abuse, age, behavioral of substance abuse have a risk to schizophrenia. Moreover, a difference between the articles and this study will a risk was likely to have schizophrenia with substance abuse behavior of dependence syndromes.

4.3 Limitation

The problem of this study is to identify the period of substance used. This data may be incorrect due to inability to identify the period of use for certain substance. In addition, the sources of data from the hospital are located in Bangkok, so the results from the association rules, who lives in Bangkok.

CHAPTER V
CONCLUSION

Schizophrenia was considered to be a multifactorial disorder which the psychoactive substance usage is one of the major factors. This study was conducted by a group of schizophrenia patients with a history of substance abuse. The study is a retrospective from the database using data mining techniques by association rule discovery with FP-Growth algorithm applied to multi-dimensional data by creating a data onto the model of 24 datasets of 4 categories, such as psychoactive substance use, duration of substance use, patient demographic data, and schizophrenia status. The resulted male of diagnosis ages range of 35 to 44 years living in Bangkok of Thailand with the use of nicotine and amphetamine abused during a period of 7 to 12 months was likely to have schizophrenia and will have a higher were likely to have schizophrenia with substance abused behavior of dependence syndromes. This study showed that the data mining technique by association rules discovery with FP-Growth algorithm is a promising tool to reveal the hidden relationships of the underlying condition with the terrific diseases.

Further, then discovered the association rules between schizophrenia and other factors, this study also highlighted the effectiveness of FP-Growth algorithm. The findings suggested that FP-Growth was a valid analytical tool to address the research questions, and had the possibility of becoming a tool the choice of the future.

The future study may focus the effect of substance abuses the permanent change of neurons, especially in dopaminergic pathways, which is the main pathologic mechanism of schizophrenia, and other conditions that probably cause schizophrenia.

REFERENCES

1 Human Rights Watch, World Report 2014-United States, 21 January 2012, available at: http://www.refworld.org/docid/52ddfdda363.html[accessed 14 September 2015]

2 Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In *ACM SIGMOD Record* (Vol. 22, No. 2, pp. 207-216). ACM.

3 Theera leelanantakit, MD. Schizophrenia with neuroleptic malignant syndrome. 2007

4 Manote Lotrakul, MD., Pramote Sukanich, MD. Ramathibodi Essential Psychiatry Third Edition. 3nd Edition Bangkok . 2012

5 Vatinee Sukmak. Mental Health and Psychiatric Nursing 2nd Edition. 2013.

6 Richards, G., Rayward-Smith, V. J., Sönksen, P. H., Carey, S., & Weng, C. (2001). Data mining for indicators of early mortality in a database of clinical records. *Artificial intelligence in medicine*, *22*(3), 215-231.

7 Kaur, H., & Wasan, S. K. (2006). Empirical study on applications of data mining techniques in healthcare. *Journal of Computer Science*, *2*(2), 194-200.

8 Kirshners, A., Parshutin, S., & Leja, M. (2012). Research on application of data mining methods to diagnosing gastric cancer. In *Advances in Data Mining. Applications and Theoretical Aspects* (pp. 24-37). Springer Berlin Heidelberg.

9 Han, J., & Kamber, M. (2006). Mining Frequent Patterns, Associations, and Correlations. *Data Mining: Concepts and Techniques (2nd ed., pp. 227-283). San Francisco, USA: Morgan Kaufmann Publishers*.

10 Linoff, G. S., & Berry, M. J. (2011). Data mining techniques: for marketing, sales, and customer relationship management. *John Wiley & Sons*.

11 IBM SPSS Modeler CRISP-DM Guide: IBM Corporation 1994, 2011

12 Azevedo, A. & Santos, M. F. (2008), KDD, SEMMA and CRISP-DM: a parallel overview., *in* Ajith Abraham, ed., 'IADIS European Conf. Data Mining' , IADIS, , pp. 182-185.

13 Tan, P.N., Steinbach, M.,& Kumar, V. (2006). *Introduction to data mining, Boston,* MA: Pearson Addison Wesley.

14 Hu, R. (2010). Medical Data Mining Based on Association Rules. *Computer and Information Science*, *3*(4), p104.

15 Sombut Pratakkunvongsa, M. N. S. (2011). The anger control program for patients with substance induced psychosis. *Journal of Mental Health of Thailand, 18*(2), 105-112.

16 Callaghan, R. C., Cunningham, J. K., Allebeck, P., Arenovich, T., Sajeev, G., Remington, G., & Kish, S. J. (2012). Methamphetamine use and schizophrenia: a population-based cohort study in California. *American Journal of Psychiatry*, *169*(4), 389-396.

17 Jena, L., & Kamila, N. K. (2014, December). A Model for Prediction of Human Depression Using Apriori Algorithm. In *Information Technology (ICIT), 2014 International Conference on* (pp. 240-244). IEEE.

18 Vanitha, K. (2011). Evaluating The Performance Of Association Rule Mining Algorithms. *Journal of Global Research in Computer Science*, 2 (6), 101-103.

19 Howes, C., Purver, M., McCabe, R., Healey, P. G., & Lavelle, M. (2012, July). Predicting adherence to treatment for schizophrenia from dialogue transcripts. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (pp. 79-83). Association for Computational Linguistics.

20 Hunyadi, D. (2011, April). Performance comparison of Apriori and FP-Growth algorithms in generating association rules. *In Proceedings of the European Computing Conference ISBN* (pp. 978-960).

21 Medhus, S., Mordal, J., Holm, B., Mørland, J., & Bramness, J. G. (2013). A comparison of symptoms and drug use between patients with methamphetamine associated psychoses and patients diagnosed with

schizophrenia in two acute psychiatric wards. *Psychiatry research*, *206* (1), 17-21.

22 Han, Jiawei, and Micheline Kamber. *Data Mining, Southeast Asia Edition: Concepts and Techniques.* Morgan kaufmann, 2006.

23 Prabhu, S. (2007). *Data mining and warehousing*. New Age International.

24 Han, J., Pei, J., & Yin, Y. (2000, May). Mining frequent patterns without candidate generation. In *ACM SIGMOD Record* (Vol. 29, No. 2, pp. 1-12). ACM.

25 Azevedo, P. J., & Jorge, A. M. (2007). Comparing rule measures for predictive association rules. In *Machine Learning: ECML 2007* (pp. 510-517). Springer Berlin Heidelberg.

APPENDICES

APPENDIX A
RESULTS OF THE ASSOCIATION RULES

FP-Growth Association Rules of 24 datasets

1. Substance between Substances

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Alcohol, Amphetamine, Opioids	Substances	0.00	0.60	7.43	2.30
2	Alcohol, Amphetamine, Opioids	Cannabinoids	0.00	0.60	24.40	2.44
3	Nicotine, Cocaine	Alcohol	0.00	0.67	1.09	1.17
4	Opioids, Volatile	Cannabinoids	0.00	0.67	27.11	2.93
5	Alcohol, Substances, Opioids	Cannabinoids	0.00	0.67	27.11	2.93
6	Alcohol, Opioids, Cannabinoids	Substances	0.00	0.67	8.26	2.76
7	Substances, Opioids, Cannabinoids	Alcohol	0.00	0.67	1.09	1.17
8	Substances, Opioids, Cannabinoids	Amphetamine	0.00	0.67	6.23	2.68
10	Alcohol, Amphetamine, Substances, Opioids	Cannabinoids	0.00	0.67	27.11	2.93
12	Alcohol, Amphetamine, Opioids, Cannabinoids	Substances	0.00	0.67	8.26	2.76
14	Substances, Volatile	Amphetamine	0.00	0.71	6.67	3.13

2. Duration between Substances

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Nicotine, Duration 67-72	Alcohol	0.00	0.70	1.14	1.30
2	Substances, Volatile	Amphetamine	0.00	0.71	6.67	3.13
3	Duration 37-42	Alcohol	0.01	0.72	1.18	1.40
4	Duration 49-54	Alcohol	0.01	0.74	1.21	1.48
5	Duration 85-90	Alcohol	0.00	0.74	1.21	1.49
6	Nicotine, Duration > 90	Alcohol	0.00	0.75	1.23	1.55
7	Amphetamine, Duration > 90	Nicotine	0.00	0.75	3.58	3.16
8	Alcohol, Amphetamine, Duration 49-54	Nicotine	0.00	0.75	3.58	3.16
9	Nicotine, Amphetamine, Duration 49-54	Alcohol	0.00	0.75	1.23	1.55
10	Alcohol, Amphetamine, Duration > 90	Nicotine	0.00	0.75	3.58	3.16
11	Nicotine, Substances, Duration > 90	Alcohol	0.00	0.75	1.23	1.55
12	Alcohol, Duration 7-12, Cannabinoids	Nicotine	0.00	0.75	3.58	3.16
13	Alcohol, Amphetamine, Duration > 90	Substances	0.00	0.75	9.29	3.68
14	Nicotine, Substances, Duration 85-90	Amphetamine	0.00	0.75	7.01	3.57
15	Duration 0-6, Amphetamine, Substances, Opioids	Alcohol	0.00	0.75	1.23	1.55
16	Duration 79-84	Alcohol	0.00	0.76	1.25	1.63
17	Duration > 90	Alcohol	0.01	0.78	1.28	1.76
18	Duration 67-72	Alcohol	0.00	0.79	1.29	1.85
19	Amphetamine, Duration 67-72	Substances	0.00	0.80	9.91	4.60
20	Amphetamine, Duration 7-12, Cannabinoids	Nicotine	0.00	0.80	3.82	3.95
21	Duration 61-66	Alcohol	0.00	0.83	1.36	2.33
22	Substances, Duration 85-90	Amphetamine	0.00	0.83	7.79	5.36
23	Substances, Duration > 90	Alcohol	0.00	0.86	1.40	2.72

3. Social data between Substances

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Diagnosis age range 45-54, Northeast region	Alcohol	0.01	0.91	1.49	4.23
2	Married, Diagnosis age range 45-54, Farmer	Alcohol	0.00	0.91	1.49	4.27
3	Diagnosis age range 45-54, Divorced, Northeast region	Alcohol	0.00	0.91	1.49	4.27
4	Male, Diagnosis age range 45-54, Divorced, Northeast region	Alcohol	0.00	0.91	1.49	4.27
5	Married, Unemployed, Female, Diagnosis age range 15-24	Amphetamine	0.00	0.91	8.49	9.82
6	Married, Diagnosis age range 45-54, Southern region	Alcohol	0.00	0.92	1.50	4.66
7	Male, Married, Diagnosis age range 45-54, Southern region	Alcohol	0.00	0.92	1.50	4.66
8	Single, Diagnosis age range 45-54, Central region, Goverment officer	Alcohol	0.00	0.92	1.50	4.66
9	Male, Diagnosis age range 45-54, Northeast region	Alcohol	0.01	0.92	1.50	4.86
10	Male, Married, Unemployed, Northeast region	Alcohol	0.00	0.92	1.50	4.86
11	Male, Single, Diagnosis age range 55-64, Trade	Alcohol	0.00	0.92	1.51	5.05
12	Male, Single, Unemployed, Diagnosis age range 45-54, Northeast region	Alcohol	0.00	0.92	1.51	5.05
13	Male, Hireling, Diagnosis age range 45-54, Central region, Divorced	Alcohol	0.00	0.92	1.51	5.05
14	Male, Bangkok, Hireling, Diagnosis age range 55-64, Divorced	Alcohol	0.00	0.93	1.51	5.25
15	Single, Unemployed, Diagnosis age range 45-54, Northeast region	Alcohol	0.00	0.93	1.52	5.44
16	Male, Married, Hireling, Diagnosis age range 45-54, Northeast region	Alcohol	0.00	0.93	1.53	5.83
17	Bangkok, Married, Female, Diagnosis age range 15-24	Amphetamine	0.00	0.93	8.72	13.39
18	Male, Hireling, Diagnosis age range 45-54, Northeast region	Alcohol	0.00	0.94	1.53	6.22
19	Male, Hireling, Diagnosis age range 35-44, Divorced	Alcohol	0.00	0.94	1.54	6.61
20	Hireling, Diagnosis age range 45-54, Northeast region	Alcohol	0.01	0.94	1.54	6.80
21	Married, Hireling, Diagnosis age range 45-54, Northeast region	Alcohol	0.00	0.94	1.54	6.99
22	Male, Married, Diagnosis age range 65-74, Northeast region	Alcohol	0.00	0.95	1.56	8.55
23	Male, Diagnosis age range 65-74, Northeast region	Alcohol	0.00	0.96	1.57	9.33
24	Male, Single, Diagnosis age range 45-54, Northeast region	Alcohol	0.00	0.96	1.57	9.71
25	Single, Diagnosis age range 45-54, Northeast region	Alcohol	0.00	0.96	1.57	10.49
26	Male, Unemployed, Diagnosis age range 45-54, Northeast region	Alcohol	0.00	0.96	1.57	10.49
27	Unemployed, Diagnosis age range 45-54, Northeast region	Alcohol	0.00	0.97	1.58	11.66

4. Behavioural disorder between Substances

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Withdrawal state, Withdrawal state with delirium, Other mental and behavioural disorders	Alcohol	0.00	0.90	1.47	3.89
2	Dependence syndrome, Withdrawal state, Withdrawal state with delirium, Other mental and behavioural disorders	Alcohol	0.00	0.90	1.47	3.89
3	Withdrawal state, Psychotic disorder	Alcohol	0.01	0.91	1.48	4.11
4	Dependence syndrome, Withdrawal state, Other mental and behavioural disorders	Alcohol	0.01	0.91	1.49	4.27
5	Dependence syndrome, Acute intoxication, Other mental and behavioural disorders	Alcohol	0.00	0.91	1.49	4.27
6	Dependence syndrome, Withdrawal state, Psychotic disorder	Alcohol	0.01	0.91	1.50	4.53
7	Dependence syndrome, Amnesic syndrome	Alcohol	0.00	0.92	1.50	4.66
8	Dependence syndrome, Withdrawal state	Alcohol	0.07	0.92	1.50	4.73
9	Withdrawal state with delirium, Other mental and behavioural disorders	Alcohol	0.00	0.93	1.52	5.44
10	Dependence syndrome, Withdrawal state with delirium, Other mental and behavioural disorders	Alcohol	0.00	0.93	1.52	5.44
11	Harmful use, Withdrawal state	Alcohol	0.01	0.93	1.52	5.70
12	Dependence syndrome, Psychotic disorder, Acute intoxication	Alcohol	0.00	0.94	1.53	6.22
13	Withdrawal state, Psychotic disorder, Withdrawal state with delirium	Alcohol	0.00	0.94	1.54	6.61
14	Dependence syndrome, Withdrawal state, Psychotic disorder, Withdrawal state with delirium	Alcohol	0.00	0.94	1.54	6.61
15	Psychotic disorder, Withdrawal state with delirium	Alcohol	0.01	0.95	1.55	7.19
16	Dependence syndrome, Harmful use, Withdrawal state	Alcohol	0.01	0.95	1.55	7.19
17	Withdrawal state with delirium	Alcohol	0.04	0.95	1.55	7.41
18	Dependence syndrome, Psychotic disorder, Withdrawal state with delirium	Alcohol	0.00	0.97	1.59	12.82
19	Dependence syndrome, Withdrawal state, Acute intoxication	Alcohol	0.01	0.97	1.59	13.60
20	Withdrawal state, Acute intoxication	Alcohol	0.01	0.97	1.59	14.77
21	Dependence syndrome, Withdrawal state with delirium	Alcohol	0.03	0.98	1.60	17.00
22	Withdrawal state, Withdrawal state with delirium	Alcohol	0.02	0.98	1.60	19.43
23	Dependence syndrome, Withdrawal state, Withdrawal state with delirium	Alcohol	0.01	0.99	1.62	33.03

5. Behavioural disorder and duration between Substances

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Withdrawal state, Duration 25-30	Alcohol	0.00	0.90	1.47	3.89
2	Withdrawal state with delirium, Duration 25-30	Alcohol	0.00	0.90	1.47	3.89
3	Other mental and behavioural disorders, Duration 37-42	Alcohol	0.00	0.90	1.47	3.89
4	Dependence syndrome, Withdrawal state, Duration 25-30	Alcohol	0.00	0.90	1.47	3.89
5	Dependence syndrome, Withdrawal state with delirium, Duration 25-30	Alcohol	0.00	0.90	1.47	3.89
6	Dependence syndrome, Other mental and behavioural disorders, Duration 37-42	Alcohol	0.00	0.90	1.47	3.89
7	Psychotic disorder, Duration 55-60	Alcohol	0.00	0.91	1.49	4.27
8	Duration 0-6, Psychotic disorder, Withdrawal state with delirium	Alcohol	0.00	0.91	1.49	4.27
9	Dependence syndrome, Psychotic disorder, Duration 55-60	Alcohol	0.00	0.91	1.49	4.27
10	Duration 0-6, Dependence syndrome, Harmful use, Psychotic disorder	Amphetamine	0.00	0.92	8.56	10.72
11	Duration 0-6, Dependence syndrome, Withdrawal state	Alcohol	0.04	0.92	1.50	4.82
12	Withdrawal state, Duration 67-72	Alcohol	0.00	0.93	1.52	5.44
13	Duration 0-6, Withdrawal state, Other mental and behavioural disorders	Alcohol	0.00	0.93	1.52	5.44
14	Dependence syndrome, Withdrawal state, Duration 37-42	Alcohol	0.00	0.93	1.52	5.44
15	Dependence syndrome, Withdrawal state, Duration 67-72	Alcohol	0.00	0.93	1.52	5.44
16	Duration 0-6, Dependence syndrome, Withdrawal state, Other mental and behavioural disorders	Alcohol	0.00	0.93	1.52	5.44
17	Withdrawal state, Duration 43-48	Alcohol	0.00	0.93	1.53	5.83
18	Duration 0-6, Withdrawal state, Psychotic disorder	Alcohol	0.00	0.95	1.56	8.16
19	Duration 0-6, Withdrawal state with delirium	Alcohol	0.03	0.96	1.56	8.84
20	Duration 0-6, Dependence syndrome, Withdrawal state with delirium	Alcohol	0.01	0.99	1.61	30.31

6. Behavioural disorder and social data between Substances

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Married, Withdrawal state, Withdrawal state with delirium	Alcohol	0.01	0.98	1.61	21.37
2	Male, Dependence syndrome, Single, Diagnosis age range 45-54, Withdrawal state	Alcohol	0.01	0.98	1.61	21.37
3	Dependence syndrome, Married, Hireling, Diagnosis age range 45-54, Withdrawal state	Alcohol	0.01	0.98	1.61	22.15
4	Bangkok, Single, Withdrawal state with delirium	Alcohol	0.01	0.98	1.61	22.54
5	Male, Married, Hireling, Acute intoxication	Alcohol	0.01	0.98	1.61	22.92
6	Male, Hireling, Withdrawal state with delirium	Alcohol	0.02	0.98	1.61	23.51
7	Male, Diagnosis age range 35-44, Withdrawal state with delirium	Alcohol	0.01	0.98	1.61	24.09
8	Male, Dependence syndrome, Bangkok, Diagnosis age range 45-54, Withdrawal state	Alcohol	0.02	0.98	1.61	24.09
9	Male, Bangkok, Hireling, Diagnosis age range 35-44, Withdrawal state	Alcohol	0.01	0.98	1.61	24.48
10	Male, Hireling, Diagnosis age range 45-54, Withdrawal state	Alcohol	0.02	0.98	1.61	25.26
11	Male, Dependence syndrome, Married, Diagnosis age range 55-64, Withdrawal state	Alcohol	0.01	0.98	1.61	25.64
12	Male, Bangkok, Married, Hireling, Diagnosis age range 45-54, Withdrawal state	Alcohol	0.01	0.98	1.61	25.64
13	Dependence syndrome, Married, Diagnosis age range 55-64, Withdrawal state	Alcohol	0.01	0.99	1.61	26.81
14	Male, Dependence syndrome, Bangkok, Hireling, Withdrawal state with delirium	Alcohol	0.01	0.99	1.61	26.81
15	Male, Married, Diagnosis age range 45-54, Withdrawal state	Alcohol	0.02	0.99	1.61	27.39
16	Dependence syndrome, Bangkok, Hireling, Withdrawal state with delirium	Alcohol	0.01	0.99	1.61	28.36
17	Male, Dependence syndrome, Bangkok, Married, Withdrawal state with delirium	Alcohol	0.01	0.99	1.61	29.14
18	Male, Dependence syndrome, Bangkok, Hireling, Diagnosis age range 45-54, Withdrawal state	Alcohol	0.01	0.99	1.61	29.92
19	Male, Dependence syndrome, Bangkok, Diagnosis age range 55-64, Withdrawal state	Alcohol	0.01	0.99	1.61	30.70
20	Male, Married, Hireling, Diagnosis age range 45-54, Withdrawal state	Alcohol	0.01	0.99	1.62	31.08
21	Dependence syndrome, Bangkok, Married, Withdrawal state with delirium	Alcohol	0.01	0.99	1.62	31.47
22	Male, Dependence syndrome, Withdrawal state with delirium	Alcohol	0.03	0.99	1.62	31.86
23	Male, Dependence syndrome, Diagnosis age range 45-54, Withdrawal state	Alcohol	0.03	0.99	1.62	31.86
24	Dependence syndrome, Withdrawal state, Withdrawal state with delirium	Alcohol	0.01	0.99	1.62	33.03
25	Male, Dependence syndrome, Hireling, Withdrawal state with delirium	Alcohol	0.01	0.99	1.62	33.03
26	Dependence syndrome, Hireling, Withdrawal state with delirium	Alcohol	0.01	0.99	1.62	34.58
27	Diagnosis age range 45-54, Withdrawal state with delirium	Alcohol	0.01	0.99	1.62	35.36
28	Male, Dependence syndrome, Hireling, Diagnosis age range 45-54, Withdrawal state	Alcohol	0.01	0.99	1.62	35.75
29	Male, Withdrawal state, Withdrawal state with delirium	Alcohol	0.01	0.99	1.62	36.52
30	Male, Dependence syndrome, Bangkok, Married, Hireling, Withdrawal state	Alcohol	0.01	0.99	1.62	37.30
31	Male, Dependence syndrome, Married, Withdrawal state with delirium	Alcohol	0.01	0.99	1.62	37.69
32	Male, Dependence syndrome, Diagnosis age range 55-64, Withdrawal state	Alcohol	0.02	0.99	1.62	39.63
33	Dependence syndrome, Married, Withdrawal state with delirium	Alcohol	0.02	0.99	1.62	40.02
34	Male, Dependence syndrome, Bangkok, Withdrawal state with delirium	Alcohol	0.02	0.99	1.62	47.79
35	Dependence syndrome, Bangkok, Withdrawal state with delirium	Alcohol	0.02	0.99	1.62	51.29

7. Behavioural disorder between Behavioural disorders

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Psychotic disorder, Other mental and behavioural disorders	Dependence syndrome	0.01	0.80	1.13	1.48
2	Psychotic disorder, Amnesic syndrome	Withdrawal state with delirium	0.00	0.80	18.08	4.78
3	Withdrawal state with delirium, Amnesic syndrome	Psychotic disorder	0.00	0.80	8.72	4.54
5	Dependence syndrome, Psychotic disorder, Amnesic syndrome	Withdrawal state with delirium	0.00	0.80	18.08	4.78
7	Dependence syndrome, Withdrawal state with delirium, Amnesic syndrome	Psychotic disorder	0.00	0.80	8.72	4.54
8	Harmful use, Withdrawal state with delirium	Dependence syndrome	0.00	0.83	1.18	1.77
9	Harmful use, Withdrawal state	Dependence syndrome	0.01	0.84	1.19	1.86
10	Psychotic disorder, Acute intoxication	Dependence syndrome	0.00	0.84	1.19	1.87
11	Withdrawal state, Withdrawal state with delirium	Dependence syndrome	0.01	0.85	1.21	1.97
12	Withdrawal state, Unspecified mental and behavioural disorder	Dependence syndrome	0.00	0.86	1.22	2.07
13	Harmful use, Psychotic disorder, Other mental and behavioural disorders	Dependence syndrome	0.00	0.87	1.23	2.21
14	Acute intoxication, Withdrawal state with delirium	Dependence syndrome	0.00	0.88	1.25	2.51
15	Psychotic disorder, Withdrawal state with delirium	Dependence syndrome	0.01	0.89	1.27	2.73
16	Withdrawal state, Acute intoxication, Withdrawal state with delirium	Dependence syndrome	0.00	0.91	1.29	3.25
17	Acute intoxication, Other mental and behavioural disorders	Dependence syndrome	0.00	0.92	1.30	3.54
18	Withdrawal state, Acute intoxication	Dependence syndrome	0.01	0.92	1.31	3.74
19	Withdrawal state, Psychotic disorder	Dependence syndrome	0.01	0.95	1.34	5.46
20	Withdrawal state, Other mental and behavioural disorders	Dependence syndrome	0.01	0.98	1.39	13.28

8. Duration between Behavioural disorders

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Harmful use, Duration 67-72	Dependence syndrome	0.00	0.90	1.28	2.95
2	Psychotic disorder, Duration 43-48	Dependence syndrome	0.00	0.90	1.28	2.95
3	Acute intoxication, Duration > 90	Dependence syndrome	0.00	0.90	1.28	2.95
4	Other mental and behavioural disorders, Duration 25-30	Dependence syndrome	0.00	0.90	1.28	2.95
5	Psychotic disorder, Duration 19-24	Dependence syndrome	0.00	0.91	1.29	3.25
6	Duration 25-30	Dependence syndrome	0.02	0.91	1.29	3.27
7	Duration 19-24	Dependence syndrome	0.02	0.91	1.29	3.31
8	Duration 85-90	Dependence syndrome	0.00	0.91	1.30	3.39
9	Other mental and behavioural disorders, Duration 13-18	Dependence syndrome	0.00	0.92	1.30	3.54
10	Harmful use, Duration 37-42	Dependence syndrome	0.00	0.92	1.31	3.84
11	Harmful use, Duration 49-54	Dependence syndrome	0.00	0.92	1.31	3.84
12	Withdrawal state, Duration 19-24	Dependence syndrome	0.00	0.93	1.31	3.98
13	Harmful use, Duration > 90	Dependence syndrome	0.00	0.93	1.32	4.13
14	Psychotic disorder, Duration 49-54	Dependence syndrome	0.00	0.93	1.32	4.13
15	Withdrawal state, Duration 37-42	Dependence syndrome	0.00	0.93	1.32	4.43
16	Withdrawal state, Duration 43-48	Dependence syndrome	0.00	0.93	1.32	4.43
17	Withdrawal state, Duration > 90	Dependence syndrome	0.00	0.93	1.32	4.43
18	Duration 37-42	Dependence syndrome	0.01	0.94	1.33	4.66
19	Duration 43-48	Dependence syndrome	0.01	0.94	1.33	4.80
20	Duration 31-36	Dependence syndrome	0.01	0.94	1.33	4.82
21	Psychotic disorder, Duration > 90	Dependence syndrome	0.00	0.94	1.34	5.02
22	Duration 55-60	Dependence syndrome	0.01	0.94	1.34	5.21
23	Withdrawal state with delirium, Duration 13-18	Dependence syndrome	0.00	0.94	1.34	5.31
24	Duration 49-54	Dependence syndrome	0.01	0.95	1.35	6.00
25	Duration 79-84	Dependence syndrome	0.00	0.95	1.35	6.20
26	Withdrawal state, Duration 49-54	Dependence syndrome	0.00	0.95	1.35	6.20

9. Social data between Behavioural disorders

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Married, Central region, Diagnosis age range 55-64, Goverment officer	Dependence syndrome	0.01	0.95	1.34	5.39
2	Male, Diagnosis age range 45-54, Withdrawal state, Withdrawal state with delirium	Dependence syndrome	0.01	0.95	1.34	5.46
3	Bangkok, Withdrawal state, Psychotic disorder	Dependence syndrome	0.01	0.95	1.34	5.51
4	Married, Psychotic disorder, Other mental and behavioural disorders	Dependence syndrome	0.00	0.95	1.34	5.61
5	Male, Diagnosis age range 45-54, Withdrawal state, Psychotic disorder	Dependence syndrome	0.00	0.95	1.34	5.61
6	Married, Withdrawal state, Psychotic disorder	Dependence syndrome	0.01	0.95	1.35	5.75
7	Diagnosis age range 45-54, Withdrawal state, Withdrawal state with delirium	Dependence syndrome	0.01	0.95	1.35	5.75
8	Hireling, Withdrawal state, Other mental and behavioural disorders	Dependence syndrome	0.00	0.95	1.35	6.20
9	Diagnosis age range 35-44, Withdrawal state, Psychotic disorder	Dependence syndrome	0.00	0.95	1.35	6.20
10	Male, Withdrawal state, Psychotic disorder	Dependence syndrome	0.01	0.95	1.35	6.30
11	Male, Married, Central region, Diagnosis age range 55-64, Goverment officer	Dependence syndrome	0.01	0.96	1.36	6.69
12	Male, Single, Withdrawal state, Psychotic disorder	Dependence syndrome	0.00	0.96	1.36	6.79
13	Male, Bangkok, Withdrawal state, Psychotic disorder	Dependence syndrome	0.01	0.96	1.36	7.23
14	Single, Withdrawal state, Psychotic disorder	Dependence syndrome	0.00	0.96	1.36	7.67
15	Male, Bangkok, Married, Withdrawal state, Psychotic disorder	Dependence syndrome	0.00	0.96	1.37	8.26
16	Male, Bangkok, Hireling, Withdrawal state, Psychotic disorder	Dependence syndrome	0.00	0.97	1.37	8.85
17	Bangkok, Hireling, Withdrawal state, Psychotic disorder	Dependence syndrome	0.00	0.97	1.37	9.44
18	Male, Married, Withdrawal state, Psychotic disorder	Dependence syndrome	0.01	0.97	1.38	10.03
19	Male, Hireling, Withdrawal state, Psychotic disorder	Dependence syndrome	0.01	0.97	1.38	10.33

10. Substance between Schizophrenia

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Nicotine, Volatile	Schizophrenia	0.01	0.75	1.11	1.30
2	Amphetamine, Sedatives of hypnotics	Schizophrenia	0.01	0.78	1.15	1.47
3	Nicotine, Alcohol	Schizophrenia	0.10	0.80	1.19	1.63
4	Nicotine, Alcohol, Amphetamine	Schizophrenia	0.02	0.82	1.21	1.79
5	Nicotine	Schizophrenia	0.35	0.85	1.26	2.15
6	Nicotine, Sedatives of hypnotics	Schizophrenia	0.01	0.86	1.27	2.28
7	Nicotine, Amphetamine	Schizophrenia	0.05	0.88	1.30	2.69
8	Nicotine, Substances	Schizophrenia	0.03	0.88	1.31	2.77
9	Nicotine, Amphetamine, Substances	Schizophrenia	0.01	0.89	1.32	2.93

11. Substance and duration between Schizophrenia

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Nicotine, Duration 7-12	Schizophrenia	0.03	0.80	1.19	1.63
2	Alcohol, Duration 37-42	Schizophrenia	0.01	0.80	1.19	1.63
3	Amphetamine, Duration 13-18	Schizophrenia	0.01	0.80	1.19	1.63
4	Nicotine, Alcohol, Duration 31-36	Schizophrenia	0.01	0.80	1.19	1.63
5	Alcohol, Duration 19-24	Schizophrenia	0.02	0.86	1.27	2.28
6	Nicotine, Alcohol, Duration>90	Schizophrenia	0.01	0.88	1.30	2.61
7	Nicotine, Duration 13-18	Schizophrenia	0.02	0.90	1.33	3.26
8	Alcohol, Duration>90	Schizophrenia	0.02	0.90	1.33	3.26
9	Nicotine, Duration>90	Schizophrenia	0.02	0.91	1.35	3.58
10	Nicotine, Duration 25-30	Schizophrenia	0.02	0.92	1.37	4.23
11	Amphetamine, Duration 7-12	Schizophrenia	0.03	0.94	1.39	5.21

12. Substance and social data between Schizophrenia

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Male, Nicotine, Goverment officer	schizophrenia	0.04	0.93	1.37	4.40
2	Male, Trade, Nicotine, Married	schizophrenia	0.02	0.93	1.38	4.56
3	Male, Trade, Nicotine, Diagnosis age range 45-54	schizophrenia	0.02	0.93	1.38	4.56
4	Bangkok, Single, Nicotine, Amphetamine	schizophrenia	0.02	0.93	1.38	4.56
5	Nicotine, Goverment officer	schizophrenia	0.05	0.93	1.38	4.72
6	Trade, Nicotine, Diagnosis age range 45-54	schizophrenia	0.02	0.93	1.38	4.88
7	Male, Single, Nicotine, Diagnosis age range 35-44	schizophrenia	0.10	0.93	1.38	4.88
8	Male, Trade, Nicotine, Amphetamine	schizophrenia	0.02	0.93	1.38	4.88
9	Single, Trade, Nicotine, Diagnosis age range 35-44	schizophrenia	0.05	0.93	1.38	4.88
10	Male, Single, Trade, Nicotine, Diagnosis age range 35-44	schizophrenia	0.05	0.93	1.38	4.88
11	Male, Single, Nicotine, Farmer, Diagnosis age range 35-44	schizophrenia	0.03	0.94	1.39	5.21
12	Bangkok, Nicotine, Diagnosis age range 55-64	schizophrenia	0.03	0.94	1.40	5.54
13	Trade, Nicotine, Married	schizophrenia	0.03	0.94	1.40	5.54
14	Bangkok, Single, Nicotine, Alcohol, Diagnosis age range 35-44	schizophrenia	0.03	0.94	1.40	5.54
15	Male, Bangkok, Single, Nicotine, Alcohol, Diagnosis age range 35-44	schizophrenia	0.03	0.94	1.40	5.54
16	Male, Bangkok, Nicotine, Goverment officer	schizophrenia	0.03	0.94	1.40	5.86
17	Male, Bangkok, Nicotine, Farmer, Diagnosis age range 35-44	schizophrenia	0.03	0.94	1.40	5.86
18	Bangkok, Nicotine, Farmer, Diagnosis age range 35-44	schizophrenia	0.03	0.95	1.40	6.19
19	Bangkok, Nicotine, Goverment officer	schizophrenia	0.03	0.95	1.41	6.51
20	Male, Single, Nicotine, Goverment officer	schizophrenia	0.04	0.95	1.41	6.84
21	Single, Nicotine, Goverment officer	schizophrenia	0.04	0.95	1.42	7.16
22	Bangkok, Single, Nicotine, Diagnosis age range 35-44	schizophrenia	0.08	0.96	1.42	7.82
23	Male, Bangkok, Single, Nicotine, Diagnosis age range 35-44	schizophrenia	0.08	0.96	1.42	7.82

13. Substance between Symptoms

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Nicotine, Alcohol, Amphetamine	Paranoid schizophrenia	0.02	0.82	1.80	3.00
2	Nicotine, Alcohol, Substances	Paranoid schizophrenia	0.01	0.83	1.83	3.27

14. Substance and duration between Symptoms

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Alcohol, Duration>90	Paranoid schizophrenia	0.01	0.80	1.76	2.73
2	Nicotine, Duration 13-18	Schizophrenia, unspecified	0.01	0.80	1.92	2.92
3	Amphetamine, Duration 13-18	Schizophrenia, unspecified	0.01	0.80	1.92	2.92
4	Nicotine, Alcohol, Duration 37-42	Paranoid schizophrenia	0.01	0.80	1.76	2.73
6	Nicotine, Duration>90	Paranoid schizophrenia	0.02	0.82	1.80	3.00
8	Nicotine, Duration 19-24	Paranoid schizophrenia	0.03	0.83	1.83	3.27
9	Nicotine, Duration 49-54	Schizophrenia, unspecified	0.01	0.83	2.00	3.50
13	Nicotine, Duration 55-60	Paranoid schizophrenia	0.01	0.89	1.95	4.91
14	Nicotine, Duration 43-48	Paranoid schizophrenia	0.02	0.90	1.98	5.45

15. Substance and social data between Symptoms

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Alcohol, Amphetamine, Central region	Paranoid schizophrenia	0.01	0.88	1.92	4.36
2	Male, Nicotine, Farmer, Central region	Paranoid schizophrenia	0.01	0.88	1.92	4.36
3	Male, Alcohol, Amphetamine, Central region	Paranoid schizophrenia	0.01	0.88	1.92	4.36
4	Single, Nicotine, Diagnosis age range 45-54, Central region	Paranoid schizophrenia	0.01	0.88	1.92	4.36
5	Bangkok, Single, Nicotine, Diagnosis age range 35-44, Goverment officer	Paranoid schizophrenia	0.01	0.88	1.92	4.36
6	Male, Bangkok, Single, Nicotine, Diagnosis age range 35-44, Goverment officer	Paranoid schizophrenia	0.01	0.88	1.92	4.36
7	Nicotine, Diagnosis age range 25-34, Substances	Schizophrenia unspecified	0.01	0.88	2.10	4.67
8	Bangkok, Nicotine, Diagnosis age range 25-34, Substances	Schizophrenia unspecified	0.01	0.88	2.10	4.67
9	Single, Nicotine, Diagnosis age range 25-34, Substances	Schizophrenia unspecified	0.01	0.88	2.10	4.67
10	Bangkok, Single, Nicotine, Diagnosis age range 25-34, Substances	Schizophrenia unspecified	0.01	0.88	2.10	4.67
11	Amphetamine, Diagnosis age range 45-54	Paranoid schizophrenia	0.01	0.89	1.95	4.91
12	Male, Amphetamine, Diagnosis age range 45-54	Paranoid schizophrenia	0.01	0.89	1.95	4.91
13	Male, Nicotine, Diagnosis age range 35-44, Central region	Paranoid schizophrenia	0.01	0.89	1.95	4.91
14	Amphetamine, Diagnosis age range 45-54	Schizophrenia unspecified	0.01	0.89	2.14	5.26
15	Male, Amphetamine, Diagnosis age range 45-54	Schizophrenia unspecified	0.01	0.89	2.14	5.26

16. Behavioural disorder between Schizophrenia

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Dependence syndrome, Harmful use	Schizophrenia	0.09	0.72	1.06	1.15
2	Dependence syndrome	Schizophrenia	0.51	0.73	1.08	1.21
3	Dependence syndrome, Other mental and behavioural disorders	Schizophrenia	0.02	0.74	1.09	1.24
4	Harmful use, Other mental and behavioural disorders	Schizophrenia	0.01	0.75	1.11	1.30
5	Psychotic disorder, Harmful use, Other mental and behavioural disorders	Schizophrenia	0.01	0.75	1.11	1.30
6	Dependence syndrome, Psychotic disorder, Unspecified mental and behavioural disorder	Schizophrenia	0.01	0.80	1.19	1.63
7	Dependence syndrome, Harmful use, Other mental and behavioural disorders	Schizophrenia	0.01	0.80	1.19	1.63
8	Dependence syndrome, Unspecified mental and behavioural disorder	Schizophrenia	0.02	0.90	1.33	3.26

17. Behavioural disorder and duration between Schizophrenia

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Psychotic disorder, Duration 55-60	Schizophrenia	0.01	0.80	1.19	1.63
2	Harmful use, Duration 25-30	Schizophrenia	0.01	0.80	1.19	1.63
3	Harmful use, Duration73-78	Schizophrenia	0.01	0.80	1.19	1.63
4	Dependence syndrome, Psychotic disorder, Duration 25-30	Schizophrenia	0.01	0.80	1.19	1.63
5	Dependence syndrome, Psychotic disorder, Duration 55-60	Schizophrenia	0.01	0.80	1.19	1.63
6	Dependence syndrome, Duration 61-66	Schizophrenia	0.01	0.83	1.24	1.95
7	Dependence syndrome, Psychotic disorder, Duration 31-36	Schizophrenia	0.01	0.83	1.24	1.95
8	Dependence syndrome, Duration 25-30	Schizophrenia	0.03	0.84	1.25	2.06
9	Dependence syndrome, Duration 37-42	Schizophrenia	0.02	0.85	1.25	2.12
10	Dependence syndrome, Duration 55-60	Schizophrenia	0.02	0.86	1.27	2.28
11	Dependence syndrome, Duration 19-24	Schizophrenia	0.05	0.87	1.29	2.44
12	Dependence syndrome, Psychotic disorder, Duration 37-42	Schizophrenia	0.01	0.88	1.30	2.61
13	Psychotic disorder, Duration 37-42	Schizophrenia	0.01	0.89	1.32	2.93
14	Harmful use, Duration 7-12	Schizophrenia	0.02	0.90	1.33	3.26
15	Dependence syndrome, Duration>90	Schizophrenia	0.02	0.92	1.37	4.23
16	Psychotic disorder, Duration 19-24	Schizophrenia	0.02	0.92	1.37	4.23

18. Behavioural disorder and social data between Schizophrenia

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Harmful use, Central region	Schizophrenia	0.02	0.80	1.19	1.63
2	Diagnosis age range 45-54, Unspecified mental and behavioural disorder	Schizophrenia	0.01	0.80	1.19	1.63
3	Male, Bangkok, Other mental and behavioural disorders	Schizophrenia	0.02	0.80	1.19	1.63

19. Behavioural disorder between Symptoms

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Harmful use, Unspecified mental and behavioural disorder	Paranoid schizophrenia	0.01	0.80	1.76	2.73
2	Harmful use, Unspecified mental and behavioural disorder	Schizophrenia-unspecified	0.01	0.80	1.82	2.81
3	Dependence syndrome, Harmful use, Unspecified mental and behavioural disorder	Paranoid schizophrenia	0.01	0.80	1.76	2.73
4	Dependence syndrome, Harmful use, Other mental and behavioural disorders	Schizophrenia-unspecified	0.01	0.80	1.82	2.81
5	Dependence syndrome, Harmful use, Unspecified mental and behavioural disorder	Schizophrenia-unspecified	0.01	0.80	1.82	2.81

20. Behavioural disorder and duration between Symptoms

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Psychotic disorder, Duration 55-60	Paranoid schizophrenia	0.01	0.80	1.76	2.73
2	Harmful use, Duration 7-12	Paranoid schizophrenia	0.01	0.80	1.76	2.73
3	Harmful use, Duration>90	Paranoid schizophrenia	0.01	0.80	1.76	2.73
4	Harmful use, Duration73-78	Paranoid schizophrenia	0.01	0.80	1.76	2.73
5	Harmful use, Duration 13-18	Schizophrenia unspecified	0.01	0.80	1.82	2.81
6	Dependence syndrome, Psychotic disorder, Duration 55-60	Paranoid schizophrenia	0.01	0.80	1.76	2.73
7	Dependence syndrome, Harmful use, Duration>90	Paranoid schizophrenia	0.01	0.80	1.76	2.73
8	Dependence syndrome, Duration>90	Paranoid schizophrenia	0.02	0.85	1.86	3.54
9	Dependence syndrome, Harmful use, Duration 7-12	Paranoid schizophrenia	0.01	0.88	1.92	4.36

21. Behavioural disorder and social data between Symptoms

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Male, Dependence syndrome, Single, Other mental and behavioural disorders	Schizophrenia unspecified	0.01	0.80	1.82	2.81

22. Mental and behavioural disorder between Schizophrenia

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	cannabinoids dependence syndrome	schizophrenia	0.02	0.80	1.19	1.63
2	Amphetamine withdrawal state	schizophrenia	0.01	0.80	1.19	1.63
3	Amphetamine unspecified mental and behavioural disorder	schizophrenia	0.01	0.80	1.19	1.63
4	Nicotine dependence syndrome, Alcohol Withdrawal state	schizophrenia	0.01	0.80	1.19	1.63
5	Nicotine dependence syndrome, cannabinoids dependence syndrome	schizophrenia	0.01	0.80	1.19	1.63
6	substances psychotic disorder, Alcohol Harmful use	schizophrenia	0.01	0.80	1.19	1.63
7	Amphetamine psychotic disorder, substances dependence syndrome	schizophrenia	0.01	0.80	1.19	1.63
8	Amphetamine dependence syndrome, cannabinoids dependence syndrome	schizophrenia	0.01	0.80	1.19	1.63
9	cannabinoids dependence syndrome, cannabinoids psychotic disorder	schizophrenia	0.01	0.80	1.19	1.63
10	Nicotine dependence syndrome, Alcohol Dependence syndrome, Alcohol Withdrawal state	schizophrenia	0.01	0.80	1.19	1.63
11	Nicotine dependence syndrome, Amphetamine psychotic disorder	schizophrenia	0.02	0.81	1.20	1.74
12	Nicotine dependence syndrome, Alcohol Dependence syndrome, Alcohol Harmful use	schizophrenia	0.02	0.81	1.20	1.74
13	Amphetamine dependence syndrome	schizophrenia	0.09	0.82	1.21	1.78
14	Nicotine unspecified mental and behavioural disorder	schizophrenia	0.01	0.83	1.24	1.95
15	substances psychotic disorder, Amphetamine psychotic disorder, Amphetamine dependence syndrome	schizophrenia	0.01	0.83	1.24	1.95
16	Amphetamine psychotic disorder, Amphetamine dependence syndrome, Amphetamine harmful use	schizophrenia	0.01	0.83	1.24	1.95
17	Nicotine dependence syndrome	schizophrenia	0.34	0.85	1.27	2.22
18	Nicotine dependence syndrome, sedatives of hypnotics dependence syndrome	schizophrenia	0.01	0.86	1.27	2.28
19	Amphetamine psychotic disorder, Amphetamine other mental and behavioural disorders	schizophrenia	0.01	0.86	1.27	2.28
20	Nicotine dependence syndrome, Amphetamine harmful use	schizophrenia	0.01	0.88	1.30	2.61
21	substances psychotic disorder, Amphetamine dependence syndrome	schizophrenia	0.01	0.88	1.30	2.61
22	Amphetamine psychotic disorder, Amphetamine harmful use	schizophrenia	0.01	0.88	1.30	2.61
23	Nicotine dependence syndrome, Amphetamine psychotic disorder, Amphetamine dependence syndrome	schizophrenia	0.01	0.88	1.30	2.61
24	Nicotine dependence syndrome, Alcohol Harmful use	schizophrenia	0.04	0.88	1.30	2.71
25	Nicotine dependence syndrome, substances psychotic disorder	schizophrenia	0.02	0.90	1.33	3.26
26	Nicotine dependence syndrome, Amphetamine dependence syndrome	schizophrenia	0.04	0.95	1.41	6.84

23. Mental and behavioural disorder and duration between Schizophrenia

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Duration 0-6 , substances dependence syndrome	schizophrenia	0.01	0.80	1.19	1.63
2	Nicotine dependence syndrome, Duration 7-12	schizophrenia	0.03	0.80	1.19	1.63
3	substances psychotic disorder, Duration 37-42	schizophrenia	0.01	0.80	1.19	1.63
4	Alcohol Harmful use, Duration 25-30	schizophrenia	0.01	0.80	1.19	1.63
5	Duration 0-6 , Nicotine dependence syndrome, Alcohol Harmful use	schizophrenia	0.01	0.80	1.19	1.63
6	Amphetamine psychotic disorder, Duration 7-12	schizophrenia	0.01	0.83	1.24	1.95
7	Alcohol Dependence syndrome, Alcohol Harmful use, Duration 19-24	schizophrenia	0.01	0.83	1.24	1.95
8	Alcohol Dependence syndrome, Duration 19-24	schizophrenia	0.02	0.85	1.25	2.12
9	Amphetamine psychotic disorder, Duration 19-24	schizophrenia	0.01	0.86	1.27	2.28
10	Alcohol Harmful use, Duration 19-24	schizophrenia	0.01	0.86	1.27	2.28
11	Nicotine dependence syndrome, Alcohol Dependence syndrome, Duration>90	schizophrenia	0.01	0.86	1.27	2.28
12	Alcohol Dependence syndrome, Duration>90	schizophrenia	0.01	0.88	1.30	2.61
13	Nicotine dependence syndrome, Duration 13-18	schizophrenia	0.02	0.90	1.33	3.26
14	Nicotine dependence syndrome, Duration>90	schizophrenia	0.02	0.91	1.35	3.58
15	Nicotine dependence syndrome, Duration 25-30	schizophrenia	0.02	0.92	1.37	4.23

24. Mental and behavioural disorder and social data between Schizophrenia

No.	Premises	Conclusion	Support	Confidence	Lift	Conviction
1	Male, Bangkok, Nicotine, dependence syndrome, Goverment officer	schizophrenia	0.03	0.94	1.40	5.86
2	Male, Bangkok, Diagnosis age range 25-34, Amphetamine, dependence syndrome	schizophrenia	0.03	0.94	1.40	5.86
3	Male, Bangkok, Single, Diagnosis age range 25-34, Amphetamine, dependence syndrome	schizophrenia	0.03	0.94	1.40	5.86
4	Male, Bangkok, Nicotine, dependence syndrome, Farmer, Diagnosis age range 35-44	schizophrenia	0.03	0.94	1.40	5.86
5	Bangkok, Trade, Diagnosis age range 25-34, Amphetamine, dependence syndrome	schizophrenia	0.03	0.95	1.40	6.19
6	Bangkok, Nicotine, dependence syndrome, Farmer, Diagnosis age range 35-44	schizophrenia	0.03	0.95	1.40	6.19
7	Bangkok, Single, Trade, Diagnosis age range 25-34, Amphetamine, dependence syndrome	schizophrenia	0.03	0.95	1.40	6.19
8	Bangkok, Nicotine, dependence syndrome, Goverment officer	schizophrenia	0.03	0.95	1.41	6.51
9	Nicotine, dependence syndrome, Amphetamine, dependence syndrome	schizophrenia	0.04	0.95	1.41	6.84
10	Male, Single, Nicotine, dependence syndrome, Goverment officer	schizophrenia	0.04	0.95	1.41	6.84
11	Bangkok, Diagnosis age range 25-34, Amphetamine, dependence syndrome	schizophrenia	0.04	0.95	1.42	7.16
12	Single, Nicotine, dependence syndrome, Goverment officer	schizophrenia	0.04	0.95	1.42	7.16
13	Bangkok, Single, Diagnosis age range 25-34, Amphetamine, dependence syndrome	schizophrenia	0.04	0.95	1.42	7.16
14	Bangkok, Single, Nicotine, dependence syndrome, Diagnosis age range 35-44	schizophrenia	0.08	0.96	1.42	7.82
15	Male, Bangkok, Single, Nicotine dependence syndrome, Diagnosis age range 35-44	schizophrenia	0.08	0.96	1.42	7.82

APPENDIX B
FP-GROWTH ALGORITHM AND DISCOVERING OF ASSOCIATION RULE IN SCHIZOPHRENIC PATIENTS WITH SUBSTANCE USE

Wichian Boonyaprapa [a], Sotarat Thammaboosadee [b], Nantawat Sitdhiraksa [c], Woraphat Rattaapha [d], Nattawut Apiwannarat [e], Patcharida Mahatchariyapong [f], Piangmas Tasneeyapant [g], Thanyathach Suwanaro [h], Yanin Suksangkharn [i]

[a,b] Information Technology Management, Faculty of Engineering, Mahidol University, Nakhorn Pathom, Thailand.

[a] wichian.boo@mahidol.ac.th, [b] sotarat.tha@mahidol.ac.th

[c-i] Department of Psychiatry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.

[c] numnam@gmail.com, [d] woraphatr@hotmail.com, [e] knot_nan@hotmail.com, [f] Pam.patcharida@gmail.com, [g] Piangmastas@gmail.com, [h] Thun_so_ave@hotmail.com, [i] Yanin.0821@gmail.com

Abstract

FP-Growth algorithm (Frequent Patterns-Growth algorithm) is an analytical tool that is spotlighted for the past few years as it is claimed to outperform Apriori, a precedent one. This study aimed to evaluate the effectiveness of FP-Growth algorithm through our study of schizophrenia. We retrieved the outpatient data of schizophrenia which are recorded in ICD-10 format. The data comprised 4 categories, namely, psychoactive substance use, duration of substance use, patient demographic data, and schizophrenia status. The 24 datasets were made from the relativity of those 4 categories, and schizophrenia. The datasets were processed using FP-Growth algorithm to determine the association rules that best represent the association of schizophrenia and the other factors. The result revealed the association between schizophrenia and substance use in 5 main features: 1) Thai men at age 35-44 years, 2)

living in Bangkok, 3) history of nicotine use, 4) history of amphetamine use for a period of 7-12 months, and 5) diagnosis of substance dependence. All of the selected association rules had an acceptable confidence level of at least 0.90. This could confirm that FP-Growth algorithm with association rule technique could demonstrate a relationship pattern of data from the large database.

Keyword: FP-Growth, Association Rule, Schizophrenia, Substance Use, Data Mining

1. Background and statement of problems

Schizophrenia is an incurable chronic mental disease that requires a continuously prolonged treatment. It brings enormously negative impacts on a patient's life, family, and social activities. Now, the worldwide is facing an epidemic of amphetamine and psychoactive substance usage. The World Health Organization (Human Right Watch, 2014) found that the numbers of regular addictive amphetamine people around the world are more than 27 million. In Thailand, more than 300,000 cases were reported as a rehabilitation treatment from more than 1,200,000 cases diagnosed with methamphetamine abuse. The number of drug abuses is dramatically increasing every year. Despite of methamphetamine cessation, risk of schizophrenia isn't decreased. Nowadays, we collect diagnostic schizophrenia data in a traditional database, for easy searching and utilization information in various analyses. This study aimed to investigate the development process, and to find a relationship between schizophrenia status and observed attributes of psychoactive substance usage. The data-mining method, i.e. the association rules discovery with FP-Growth algorithm, was performed in order to identify this association. Moreover, the final model might be applied to determine and predict the cause of schizophrenia in the future. This paper was organized as follows: section 1(this, section), section 2 presented reviews the existing literature in the schizophrenia, psychosis and data-mining techniques, section 3 presented the methodology of the study process, and section 4 presented experimental results and discussion. Finally, the conclusion is summarized.

2. Related works

Schizophrenia (Sombut, 2011 and Somjit, 2014) is one type of psychiatric disorders required medical treatment. An onset of schizophrenia is usually in adolescent which etiology is not clear. Symptoms of schizophrenia could be classified into three main categories: (1) positive symptoms, (2) negative symptoms, and (3) cognitive symptoms. The positive symptoms are included hallucinations, delusions, thought disorders, and movement disorders. The negative symptoms are included problem with motivation, social withdrawal, and diminished affective responsiveness, speech and movement. The cognitive symptoms are poor executive function, impair focus and attention problem, and impair working memory. Furthermore, schizophrenia has high recurrent rate and several psychological co-morbidities. Schizophrenia can be caused by several factors. Therefore, it is necessary to identify the actual cause of schizophrenia in each patient in order to provide the proper treatment for each patient.

An advanced information technology, such as data mining techniques, provided the opportunity to analyze and to predict the cause of this psychiatric disorder. Callaghan *et al*. (Callaghan, 2012) performed the population-based cohort study aimed to investigate the relationships of schizophrenia with the psychoactive substance abuse. The risk of schizophrenia in the methamphetamine abuse group was significantly higher than the appendicitis group (hazard ratio = 9.37). The data mining technique, i.e. an applied Apriori algorithm, could be used to uncover the level of illness from the large database of depressed people (Jena, 2014). There are several ways to perform the association rule discovery, e.g. a FT-Growth algorithm, an Apriori algorithm, etc. Comparatively, the FP-growth algorithm outperforms the Appriori algorithm (Vanitha, 2011) for mining a huge number of finding rule. The FP-growth Algorithm is efficient and scalable for identifying the complete set of frequent patterns. Two data mining techniques, i.e. support vector machines and decision tree, were used to predict schizophrenia treatments (Howes, 2012). The decision tree technique demonstrated the better accuracy (approximately 90%) comparable to the support vector machine technique. The predictive performance of the FP-Growth algorithm and the Apriori algorithm has been shown to be congruent (Hunyadi, 2011) a study of acute methamphetamine-induced psychosis and acute negative-methamphetamine schizophrenia patients did not find the statistically significant

relationship between the blood methamphetamine concentrations and the severity of psychotic symptoms (Medhus, 2013). However, the onset of schizophrenia might require the duration of psychoactive substance use.

3. Methodology

In this study, we applied the association rule technique with FP-Growth algorithms to the non identifiable schizophrenic patient data. The demographic data, the history of substance use, duration of usage and behavioral disorder were extracted from the database systems with the following criteria.

Inclusion criteria

1. The medical record of the patients who had the complete information in 4 categories we made.

2. The medical record of the schizophrenic patients and the patients who had substance related disorder who were diagnosed by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria and were recorded into the database by the International Classification of Diseases, Tenth Edition (ICD-10).

Exclusion criteria

None

The Overall Process

Data mining (Han, 2006 and Linoff, 2011) is the data analysis which aims to determine the unknown patterns or rules in large database. In this study, FP-Growth algorithm was performed in order to determine the association rule of schizophrenia and the psychoactive substances use. We processed the data analysis by (1) data-preprocessing, (2) data dimensionzing, (3) modeling, and (4) model evaluation, as shown in the Fig. 1.

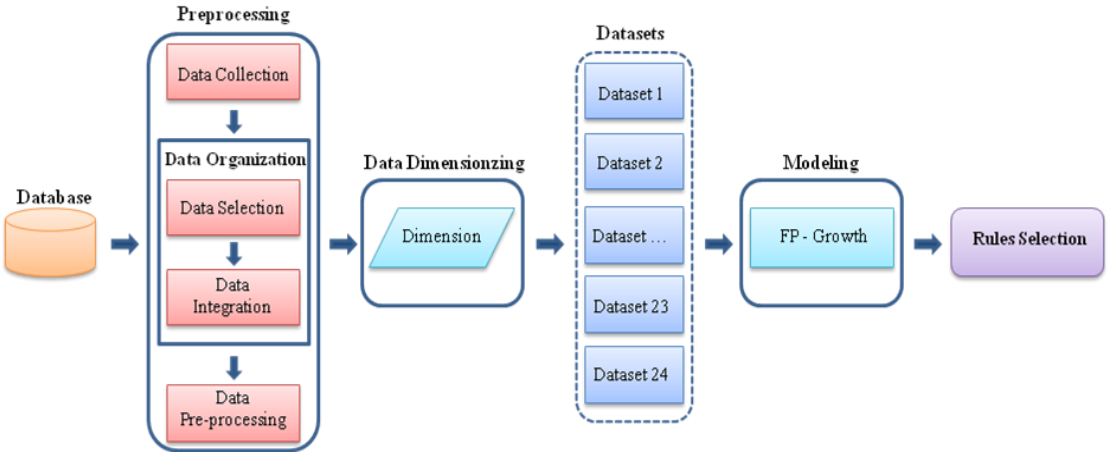


Fig. 1: The overall process

Preprocessing

(1.) Data collection

The dataset was collected from the psychiatric outpatient department clinic, including the record of schizophrenia (F20) and other mental and behavioral disorders due to the psychoactive substance usage (F10-19). The diseases were clinically diagnosed according to the diagnostic criteria of DSM-IV, and classified into the database according to the ICD-10 standard codes.

(2.) Data organization

The schizophrenia and other mental and behavioral disorder datasets consisted of the following attributes: ICD-10 code, city, sex (male or female), age at diagnosis, marital status (single, married, or divorced), and substance use (yes or no). The dataset was organized by combining the schizophrenia and other mental and behavioral disorder into the same format in order to select the patients who have them concurrently. The missing data such as the lack of any attribute data would be excluded from the dataset.

Data dimensionizing

This step we grouped the data into 24 datasets based on the combination of the following categories: psychoactive substance use, duration of the substance use, patient demographic data, and schizophrenia status. We also subcategorized the psychoactive substance use patients into 2 attributes. One was the type of substance the patients used; another was the symptoms, the behavioral disorder or both mental and behavioral disorders which were the consequence from substances use. We

selected 24 combination types which related to the demonstration of the factors associated with Schizophrenia. Our 12 datasets were composed of the data from 2 categories and another 12 datasets were composed of data from 3 categories. All 24 datasets, we selected were shown in Table 1.

Table 1: Dataset descriptions

Dataset No.	Description	Dataset No.	Description
1	Substance between Substance	13	Substance between Symptoms
2	Duration between Substance	14	Substance and duration between Symptoms
3	demographic data between Substance	15	Substance and demographic data between Symptoms
4	Behavioral disorder between Substance	16	Behavioral disorder between Schizophrenia
5	Behavioral disorder and duration between Substance	17	Behavioral disorder and duration between Schizophrenia
6	Behavioral disorder and demographic data between Substance	18	Behavioral disorder and social data between Schizophrenia
7	Behavioral disorder between Behavioral disorder	19	Behavioral disorder between Symptoms
8	Duration between Behavioral disorder	20	Behavioral disorder and duration between Symptoms
9	demographic data between Behavioral disorder	21	Behavioral disorder and demographic data between Symptoms
10	Substance between Schizophrenia	22	Mental and behavioral disorder between Schizophrenia
11	Substance and duration between Schizophrenia	23	Mental and behavioral disorder and duration between Schizophrenia
12	Substance and demographic data between Schizophrenia	24	Mental and behavioral disorder and demographic data between Schizophrenia

Modeling

In order to demonstrate the association between schizophrenia and psychoactive substance usage, this study designed to perform the association rule discovery (Han, 2006 and Prabhu, 2007). After we got the association between any attributes, the psychiatrist opinions for those associations still needed to confirm the reasonable of the discovery. To demonstrate the magnitude of the association, we applied the FP-Growth algorithm (Han, 2000), one of the association rule discovery methods to our dataset. Initially, we started with encoding the input dataset into a compact data structure called a frequent pattern tree (FP-tree).

Model Evaluation

The related parameters of association rule discovery measured by Support, Confidence, Lift and Conviction. Support is the proportion of transactions containing the set X. Confidence is the proportion of transaction containing X which also contains Y. Lift is the ratio of the observed support of that expected if X and Y were independent. A conviction can be interpreted as the ratio of the expected frequency that X occurs without Y.

4. Experimental Result and Discussion

Result

None of the subject was withdrawn from the study. The data of the subjects were analyzed in 24 datasets. Each dataset contained its specific description, which was characterized by 4 categories as shown in the previous section (Table 1). The association rules with the highest confidence level are shown in Table 2, together with its lift, and conviction. In case of equal confidence level, lift and, then, conviction were considered. Confidence, lift, and conviction are the values of probability, measure of rule, and belief of rule respectively.

Table 2: Summary of the average precision results of all experimented datasets

Dataset No.	Association Rules	Confidence	Lift	Conviction
1	Substances, Volatile → Amphetamine	0.71	6.67	3.13
2	Substances, Duration > 90 → Alcohol	0.86	1.40	2.72
3	Unemployed, Diagnosis age range, Northeast	0.97	1.58	11.66

Dataset No.	Association Rules	Confidence	Lift	Conviction
	region → Alcohol			
4	Dependence syndrome, Withdrawal state, Withdrawal state with delirium → Alcohol	0.99	1.62	33.03
5	Duration 0-6, Dependence syndrome, Withdrawal state with delirium → Alcohol	0.99	1.61	30.31
6	Male, Dependence syndrome, Bangkok, Withdrawal state with delirium → Alcohol	0.99	1.62	47.79
7	Withdrawal state, Other mental and behavioral disorders → Dependence syndrome	0.98	1.39	13.28
8	Withdrawal state, Duration 49-54 → Dependence syndrome	0.95	1.35	6.20
9	Male, Hireling, Withdrawal state, Psychotic disorder → Dependence syndrome	0.97	1.38	10.33
10	Nicotine, Amphetamine, Substances → Schizophrenia	0.89	1.32	2.93
11	Amphetamine, Duration 7-12 → Schizophrenia	0.94	1.39	5.21
12	Male, Bangkok, Single, Nicotine, Diagnosis age range 35-44 → Schizophrenia	0.96	1.42	7.82
13	Nicotine, Alcohol, Substances → Paranoid schizophrenia	0.83	1.83	3.27
14	Nicotine, Duration 43-48 → Paranoid schizophrenia	0.90	1.98	5.45
15	Male, Amphetamine, Diagnosis age range 45-54 → Schizophrenia unspecified	0.89	2.14	5.26
16	Dependence syndrome, Unspecified mental and behavioral disorder → Schizophrenia	0.90	1.33	3.26
17	Psychotic disorder, Duration 19-24 → Schizophrenia	0.92	1.37	4.23
18	Male, Bangkok, Other mental and behavioral disorders → Schizophrenia	0.80	1.19	1.63
19	Dependence syndrome, Harmful use, Unspecified mental and behavioral disorder → Schizophrenia-unspecified	0.80	1.82	2.81
20	Dependence syndrome, Harmful use, Duration 7-12 →Paranoid schizophrenia	0.88	1.92	4.36
21	Male, Dependence syndrome, Single, Other mental	0.80	1.82	2.81

Dataset No.	Association Rules	Confidence	Lift	Conviction
	and behavioral disorders → Schizophrenia unspecified			
22	Nicotine dependence syndrome, Amphetamine dependence syndrome → Schizophrenia	0.95	1.41	6.84
23	Nicotine dependence syndrome, Duration 25-30 → Schizophrenia	0.92	1.37	4.23
24	Male, Bangkok, Single, Nicotine dependence syndrome, Diagnosis age range 35-44 → Schizophrenia	0.96	1.42	7.82

Table 3: Selected rules with schizophrenia as conclusion

Dataset No.	Association Rules	Confidence	Lift	Conviction
11	Amphetamine, Duration 7-12 → Schizophrenia	0.94	1.39	5.21
12	Male, Bangkok, Single, Nicotine, Diagnosis age range 35-44 → Schizophrenia	0.96	1.42	7.82
22	Nicotine dependence syndrome, Amphetamine dependence syndrome → Schizophrenia	0.95	1.41	6.84
24	Male, Bangkok, Single, Nicotine dependence syndrome, Diagnosis age range 35-44 → Schizophrenia	0.96	1.42	7.82

Of the 24 datasets, only the datasets with the consequence of "schizophrenia" in the association rule were selected to represent the results of the study (Table 3). From the dataset No.11, 94% of the patients diagnosed with amphetamine use in the duration of 7 to 12 months were likely to have schizophrenia with the measure of rule 1.39, and with the strong belief of rule 5.21. From the dataset No.12, 96% of the single male patients in Bangkok diagnosed with nicotine use in diagnosis at the age of 35 to 44 years old were likely to have schizophrenia with the measure of rule 1.42, and with the strong belief of rule 7.82. From the dataset No.22, 95% of the patients diagnosed with nicotine dependence, and amphetamine dependence were likely to have schizophrenia with the measure of rule 1.41, and with the strong belief of rule 6.84. From the dataset No.24, 96% of the single male patients

in Bangkok diagnosed with nicotine dependence in the age of 35 to 44 years old were likely to have schizophrenia with the measure of rule 1.42, and with the strong belief of rule 7.82. To put the results together, the single male in Bangkok diagnosed with nicotine and/or amphetamine use and/or dependence at the age of 35 to 44 years old with the use of a period of 7 to 12 months was at associated with having schizophrenia.

Discussion

This study demonstrated the validity of FP-Growth algorithm as an analytical tool for research as well as of Apriori. The result of association rules showed the acceptable confidence level with the value of at least 0.70 for overall association rules and of at least 0.90 for the selected association rules. This result was also corresponding to the previous finding (Vanitha, 2011), which approved the validity of the FP - growth algorithm in comparison with Apriori algorithm.

The problem of the research was on identifying the duration of substance use, which was usually inaccurate due to inability to recall or the psychological symptoms during the substance use. Moreover, substance use and schizophrenia were often recorded by medical personnel in the same way because of the similarity in their presentation and diagnosis. The further study may focus the effect of substance use on the permanent change of neurons, especially in dopaminergic pathway, which is the main pathologic mechanism of schizophrenia, and other conditions that probably cause schizophrenia.

5. Conclusion

Further than discovering the association rules between schizophrenia and other factors, this study also highlighted the effectiveness of FP-Growth algorithm. The findings suggested that FP-Growth was a valid analytical tool to address the research questions, and had the possibility of becoming a tool of choice in the future.

6. References

Callaghan, R. C., Cunningham, J. K., Allebeck, P., Arenovich, T., Sajeev, G., Remington, G., & Kish, S. J. (2012). Methamphetamine use and

schizophrenia: a population-based cohort study in California. *American Journal of Psychiatry*, *169*(4), 389-396.

Han, J., & Kamber, M. (2006). Mining Frequent Patterns, Associations, and Correlations. *Data Mining: Concepts and Techniques (2nd ed., pp. 227-283). San Francisco, USA: Morgan Kaufmann Publishers*.

Han, J., Pei, J., & Yin, Y. (2000, May). Mining frequent patterns without candidate generation. In *ACM SIGMOD Record* (Vol. 29, No. 2, pp. 1-12). ACM.

Han, Jiawei, and Micheline Kamber. *Data Mining, Southeast Asia Edition: Concepts and Techniques*. Morgan kaufmann, 2006.

Howes, C., Purver, M., McCabe, R., Healey, P. G., & Lavelle, M. (2012, July). Predicting adherence to treatment for schizophrenia from dialogue transcripts. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (pp. 79-83). Association for Computational Linguistics.

Human Rights Watch, World Report 2014 - United States, 21 January 2014, available at: http://www.refworld.org/docid/52dfdda363.html [accessed 14 September 2015].

Hunyadi, D. (2011, April). Performance comparison of Apriori and FP-Growth algorithms in generating association rules. *In Proceedings of the European Computing Conference ISBN* (pp. 978-960).

Jena, L., & Kamila, N. K. (2014, December). A Model for Prediction of Human Depression Using Apriori Algorithm. In *Information Technology (ICIT), 2014 International Conference on* (pp. 240-244). IEEE.

Linoff, G. S., & Berry, M. J. (2011). Data mining techniques: for marketing, sales, and customer relationship management. *John Wiley & Sons*.

Maneeganol, S., Arunpongpaisal, S., Pimpanit, V., Jarassaeng, N., & Boontoch, K. (2014). Prevalence of Substance Use in Psychiatric Outpatient Clinic at Srinagarind Hospital. *J Psychiatr Assoc Thailand, 59 (4),* 371-380.

Medhus, S., Mordal, J., Holm, B., Mørland, J., & Bramness, J. G. (2013). A comparison of symptoms and drug use between patients with methamphetamine associated psychoses and patients diagnosed with

schizophrenia in two acute psychiatric wards. *Psychiatry research, 206* (1), 17-21.

Prabhu, S. (2007). *Data mining and warehousing*. New Age International.

Sombut Pratakkunvongsa, M. N. S. (2011). The anger control program for patients with substance induced psychosis. *Journal of Mental Health of Thailand, 18*(2), 105-112.

Vanitha, K. (2011). Evaluating The Performance Of Association Rule Mining Algorithms. *Journal of Global Research in Computer Science*, 2 (6), 101-103.

APPENDIX C
CERTIFICATE OF APPROVAL FOR MULTICENTER RESEARCH

COA. No. Si 223/2015

Certificate of Approval for Multicenter Research
Mahidol University

Title of Project : DEVELOPMENT OF SCHIZOPHRENIA PATIENTS CLASSIFICATION SYSTEM BASED ON MEDICAL RECORD OF AMPHETAMINES AND OTHER ABUSING SUBSTANCES USAGE

Protocol number : 070/2558(EC1)

Principal Investigator : Mr. Wichian Boonyaprapa / Faculty of Engineering, Mahidol University

Co-investigator(s) :

1) Sotarat Thammaboosadee, Ph.D. / Faculty of Engineering, Mahidol University
2) Assist. Prof. Nantawat Sitdhiraksa, MD. / Faculty of Medicine Siriraj Hospital, Mahidol University

Research site : Faculty of Medicine Siriraj Hospital, Mahidol University

Approval includes:

1) SIRB submission form
2) Protocol
3) Case Record Form
4) Curriculum Vitae

Institutional Review Boards in Mahidol University are in full compliance with International Guidelines for Human Research Protection such as Declaration of Helsinki, The Belmont Report, CIOMS Guidelines and the International Conference on Harmonization in Good Clinical Practice (ICH-GCP)

Date of Approval: April 24, 2015

Date of Expiration: April 23, 2016

Signature of Chair: .....

(Prof. Jarupim Soongswang, M.D.)

Address of IRB Siriraj Insitutional Review Board. Faculty of Medicine Siriraj Hospital, Mahidol University. His Majesty the King's 80th Birthday Anniversary 5th December 2007 Building 2nd floor Room 210 2 Wang Lang Road Bangkoknoi, Bangkok 10700, Thailand Tel. 0 2419 2667 - 72 FAX : 0 2411 0162 E-mail : siethics@ mahidol.ac.th

All MU-IRB Approved Investigators must comply with the followings:

1) Conduct the research according to the approved protocol.
2) Conduct the informed consent process without coercion or undue influence, and provide the potential subject sufficient opportunity to consider whether or not to participate, using the approved document.
3) Report to MU-IRB all of adverse event of any study subject any new information that may adversely affect the safety of the subjects or the conduct of the trial.
4) Obtain approval of any changes in research activity and informed research subjects about the change for their considerations to continuing their participations in the study.
5) Provide MU-IRB the progress report of the research annually or when requested.
6) Provide MU-IRB the closeout report when completed the study procedures.

BIOGRAPHY

NAME	Mr. Wichian Boonyaprapa
DATE OF BIRTH	18 August 1979
PLACE OF BIRTH	Phisanulok, Thailand
INSTITUTIONS ATTENDED	Thonburi College of Technology, 2004-2006 Bachelor of Sicence (Information Technology) Mahidol University, 2013-2015 Master of Science (Information Technology Management**)**
RESEARCH GRANT	This thesis is partially supported by Graduate Studies of Mahidol University Alumni Association
HOME ADDRESS	618/748 Bangbumru, Bangplat, Bangkok 10700 Tel 098-269-8393 Email: wichian.boo@mahidol.ac.th
PUBLICATION / PRESENTATION	**Boonyaprapa, W.,** Thammaboosadee , S., Sitdhiraksa, N., Ratta-apha , W., Apiwannarat, N., Mahatchariyapong, P., Tasneeyapant, P., and Suwanaro, T., FP-Growth Algorithm and Discovering of Association Rule in Schizophrenic Patients with Substance Use., Seoul International Conference on Engineering and Applied Sciences (SICEAS), South Korea, 5-7 January. 2016