Discovering Co-Occurring Medical Complications For Cocaine Users In African American Diabetic Kidney Patients: A Text Mining Method

Yong Mi Kim*, Pranay Kathuria¹ and Dursun Delen²
¹University of Oklahoma, School of Library, USA
²Oklahoma State University, USA
Received: August 31, 2018; Published: September 14, 2018
*Corresponding author: Yong Mi Kim, School of Library and Information Studies, University of Oklahoma, Schusterman Center, Tulsa, OK, 74135, USA

Abstract
The use of cocaine exacerbates kidney disease, and African Americans are more likely to use cocaine. This study explored the variety of medical problems African American diabetic kidney patients experience
a) If they use cocaine,
b) If they use both cocaine and tobacco, and
c) Further explores how these medical problems are associated with high systolic and diastolic blood pressures (SBP and DBP).

Data was procured from Cerner HealthFacts, and an apriori machine learning technique was used. This study found that cocaine users’ most frequently co-occurring issue is high SBP followed by cough and high prothrombin. While coughing, SBP and respiratory problems become prominent issues with the use of both cocaine and tobacco. Lastly, the impact of SBP and DBP are varied; elevated SBP is highly associated with low white blood cell and high BNP, while DBP is highly correlated with being overweight, respiratory problems and shortness of breath.


Introduction
Kidney diseases are one of the most exemplary cases of health disparities [1]. While high blood pressure and diabetes are partially responsible for this disparity [1], higher rates of cocaine and tobacco usage among African Americans have also driven such disparity [2-4]. To put this disparity into perspective: African Americans make up only 12.2% of the population but account for 37% of cocaine users; meaning that they are 3.5 times more likely than Caucasians to be regular cocaine users, additionally men were more likely than women to be current users of cocaine (1.0 vs. 0.3 percent) [5]. The use of cocaine results in a variety of medical problems such as kidney disease, myocardial fibrosis, high creatinine, cardiovascular issues, elevated blood pressure, and respiratory complications [6-12]. Numerous studies consistently reported high blood pressure as an issue [4,11,13] especially among African Americans [7,14,15]. Based on predominantly African American populations, a study found that cocaine users’ kidney functions declined 3.0 to 3.9 times greater than non-cocaine users [12]. As such it is not surprising that the prevalence of cocaine use among patients with hypertensive end-stage kidney disease (ESKD) is about threefold greater than non-cocaine users, and it is even higher among African Americans [16].

Cocaine use increases the urge to smoke, and thus cocaine users often use tobacco [13,17]. Furthermore, acute exposure to transdermal nicotine patches enhanced the cue for craving cocaine [18], which reinforces multiple substance abuse. Consequently, 9 in 10 people who suffer from drug addiction begin to smoke, drink and/or use other drugs [19]. This finding is confirmed by another study that 100% (20/20) of the participants used tobacco, and 16/20 (80%) had used both tobacco and marijuana prior to cocaine [20]. Because of cocaine users’ tendency to also use tobacco, scholars have explored the side effects of these two drugs. They found respiratory issues, cardiovascular diseases, increased heart rates, elevated blood pressure, various cancers, proteinuria, high microalbuminuria, and chronic kidney diseases.
Kidney Patients: A Text Mining Method

(CKD) [21-23]. Smoking especially accelerates the progression of CKD in diabetic patients [21] and is linked to high blood pressure (BP) [24]. Collectively, these reviewed studies above consistently reported positive correlations between the use of cocaine and high BP; however, it is not clear whether systolic BP (SBP), diastolic BP (DBP), or both BPs are elevated. Recent scholars claimed that SBP measure is more closely associated than DBP with development of kidney disease and cardiovascular diseases among type 2 diabetic (T2D) patients [25-27]. SBP is the main factor in determining hypertension [28], and achieved DBP is not correlated with progression of kidney disease [27].

Another report shows that 95% of patients with uncontrolled hypertension had elevated SBP levels, while DBP was elevated in only 50% of patients [29]. Taken together, those findings are pointing to SBP as a better predictor for ESKD than DBP [28, 30-33]. Although existing studies have improved our understanding of the relationship between cocaine use and its consequential medical problems, they have some problems. First, although the purpose of studies is to discover the consequential medical problems from cocaine, those studies didn’t subset only cocaine users by including tobacco users, which resulted in confounding effects within the finding. This may be the reason that consequential medical problems between sole cocaine use and use in combination with tobacco are similar. Second, studies consistently report that cocaine elevates BP as a whole. On the other hand, some other research shows that SBP is a more efficient predictor for ESKD than DBP. Again, since tobacco users are in the sample, this finding may have confounding effects. Furthermore, SBP and DBP may be differently associated with various medical problems caused by cocaine and tobacco use. Therefore, the purpose of this research is to investigate three queries using a machine learning technique. First, what kinds of co-occurring medical problems do diabetic kidney patients have if they use only cocaine? Second, what kinds of co-occurring medical problems do diabetic kidney patients have if they use both cocaine and tobacco? Third, how do these co-occurring medical problems caused by the combination of cocaine and tobacco uses associate differently with high SBP and DBP? Because African Americans are more likely to use cocaine, the focus of this study will be given to this population group.

Methods

Data and Data Processing

The data used in the study was obtained from the Cerner HealthFacts® Data Warehouse (CHFDW). The data extraction was based on the International Classification of Diseases (ICD) 9 encoding. The CHFDW is a completely Health Insurance Portability and Accountability Act (HIPAA) compliant data repository. Because this dataset is already collected by a third party, Institutional Review Board (IRB) for the Protection of Human Subjects has determined this dataset is already collected by a third party, Institutional Review Board (IRB) for the Protection of Human Subjects has determined that this study did not meet the criteria for human subject research.

Because the purpose of the paper is to discover frequently appearing co-occurring medical problems among African American diabetic kidney patients who use cocaine, T2D patients were extracted based on International Classification of Diseases, Ninth Revision (ICD-9, 250.00-250.93). They were then further filtered by those who were outpatients, African Americans, kidney patients (GFR stages between 3 and 5 including kidney patients), and cocaine users. The filtered patient records are as follows: total T2D records were 1,038,499, all kidney patients were 25,480, African American kidney outpatients and cocaine users were 439. The process of the data extraction is shown in Figure 1. Cocaine use is based on the urine test, note that the dataset combines tobacco and smoking, we used the term “tobacco.” Overweight included overweight, obesity and extreme obesity because of low counts of obesity and extreme obesity. Cough was combined to include cough, cough more than 3 weeks, and cough & deep breath. It is possible that a patient could have all three symptoms.

Analytical Strategy

Apriori was chosen as the analytical strategy for this research, as this machine-learning technique can efficiently discover frequently co-occurring medical problems. Apriori is an unsupervised machine-learning algorithm that searches all the variables in the database and retrieves only frequently appearing itemsets. As such, it does not require prior knowledge or pre-selected variables but uncover frequently appearing medical comorbidities with cocaine use. As such, this machine learning technique is ideal to discover new knowledge. Note that frequently appearing itemsets can be subjective, and thus this technique allows researchers to set the selection criteria using support, confidence, and lift values as well as use them as research finding validation: support is the fraction of transactions that contain an itemset; confidence is an accuracy of the item likely to be in the transaction; lift is a likelihood of XX and YY occur together than expected if they were statistically independent. Lift greater than 1 indicates complementary items and less than 1 indicates substitutes. Because these criteria are used to limit findings, they can control the exponential growth of the finding, and thus this method is popular among data scientists. Because this analytical strategy is widely used in marketing, it is often referred to as market basket analysis. In marketing, it is to discover what item a customer is likely to buy together; if the customer purchases milk, for example. A deeper question is what if the customer purchases milk and bread together; what item is the customer most likely to purchase? Applying this concept to this study: if a diabetic kidney patient uses cocaine, what medical problems are they likely to have, what if the patient uses both cocaine and tobacco?
Since this dataset has 3,194 medical problems, support is set to 0.000035. Confidence measures how confident it is once certain medical problems occur with the use of cocaine, this study is set to 100%. Using these criteria, the machine learning technique used 1,013 frequently appearing medical problems out of 3,194, and discovered 2,203 rules for the male group, while zero rules were discovered for the female group. It may be because the use of cocaine is higher among African American men than women[12,34]. This finding is consistent with another study that reported about 76% (37 of 49 subjects) of cocaine users were men while only 24% were women[4]. Table 1 shows five examples of the discovered results. The first rule shows that cocaine users have the characteristics of tobacco, overweight, high creatinine, high SBP, and respiratory problems. The lhs (left-hand side) and the rhs (right-hand-side) are simply a correlation, not a causation, meaning that those six medical problems co-occur. Because the lhs side variables include tobacco and overweight, it makes more sense to interpret if diabetic kidney patients are overweight, and use both cocaine and tobacco, the co-occurring medical problems are high creatinine, high SBP, and high respiratory issues. The support was 0.0000376, meaning that a chance to co-occur the set of these six medical problems out of 1013 medical problems is 0.0000376, but the probability is 100%, and the possibility for these six items to co-occur together is 4,285 times higher than the possibility of them occurring independently.

Table 1: Co-occurrence of medical problems among cocaine users.

<table>
<thead>
<tr>
<th>lh's</th>
<th>rhs</th>
<th>Support</th>
<th>Confidence</th>
<th>Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>[tobacco, overweight, high creatinine, high SBP, high respiratory] =&gt;</td>
<td>[cancer]</td>
<td>3.76e-05</td>
<td>1.0</td>
<td>4285</td>
</tr>
<tr>
<td>[high BNP, high SBP, low white blood cell, fever] =&gt;</td>
<td>[cancer]</td>
<td>3.73e-05</td>
<td>1.0</td>
<td>4285</td>
</tr>
<tr>
<td>[night sweats, high DBP, high respiratory, low white cell, fever] =&gt;</td>
<td>[cancer]</td>
<td>3.72e-05</td>
<td>1.0</td>
<td>4285</td>
</tr>
<tr>
<td>[cough, alcohol, high respiratory, fever] =&gt;</td>
<td>[cancer]</td>
<td>5.64e-05</td>
<td>1.0</td>
<td>4285</td>
</tr>
<tr>
<td>[fever, low white blood cell, overweight, shortness of breath] =&gt;</td>
<td>[cancer]</td>
<td>3.72e-05</td>
<td>1.0</td>
<td>4285</td>
</tr>
</tbody>
</table>

Results

Co-occurring medical problems among cocaine without tobacco

In order to address potential threats of confounding effects on the findings from including cocaine and tobacco users in the same analysis, we only included sole cocaine users in the first analysis and discovered 2,203 rules. If all these rules are reported individually, one cannot find systematic patterns of co-occurring medical problems, instead it becomes meaningless. Because the purpose is to discover frequently co-occurring medical problems amid cocaine users, the most frequently co-appearing set of medical problems are exclusively filtered while removing medical problems appearing below 10%.

Consistent with existing studies, this research also found high creatinine, cardiovascular issues, elevated blood pressure, and respiratory complications [6-12]. Newly discovered findings are that cocaine use elevates SBP more than DBP. In fact, high SBP is the highest medical problem among cocaine users [34]. Although this dataset has been filtered to kidney patients only, the high BNP problems occur because the use of cocaine can damage the heart as well [35]. Additional noteworthy findings are cough, high prothrombin, low red blood cell count (RBC), night sweats, fever, low white blood cell count (WBC), and low neutrophil. It seems that low RBC is from alcohol rather than from cocaine use because chronic alcohol consumption leads to deficiency of various vitamins, and one of them is folate deficiency, which disrupts the process of hematopoiesis and which in turn leads to low RBC [36,37]. Deficiency of folate leads to anemia, which is again common among alcoholics [38]. Although high occurrences of coughing, high respiratory problem, and respiratory problem with cocaine use, sparse research has done to investigate those relationships. Although studies show that cocaine and tobacco users have asthma and chronic obstructive pulmonary disease (COPD)20 and respiratory problems [39] those findings include tobacco users. The finding further shows that being over-weighted patients are more likely to use cocaine and alcohol.

Co-Occurring Medical Problems From Both Tobacco and Cocaine Uses

Tobacco use appeared 24.19% among cocaine users in this study. In this section, we further drilled down the data to discover the complications from both cocaine and tobacco use. The questions asked are: "what if diabetic kidney patients use both cocaine and tobacco, what kind of medical problems are they likely to have?" Based on these criteria, 703 rules are discovered. Figure 2 shows that high respiratory problems become the number one medical problem (36.27%) from 4th (21.80%) in Figure 3, and high BNP-B also jumped to 2nd from 6th in Figure 2. Clearly, tobacco use exacerbates respiratory problems as well as damages the heart. This finding also shows that SBP is elevated more than DBP. It is interesting though that when patients use tobacco, they tend to have high WBC to fight against the inflammation and damage caused by tobacco; however, this study found low WBC in conjunction with the use of tobacco and cocaine, which is paradoxical; however, a
Another notable finding is that overweight and alcohol have increased from Figure 2, suggesting that overweight diabetic kidney patients are likely to use multiple substances. Unexpectedly, cardiopulmonary has dropped to 9th (14.94%) from 5th (21%) in Figure 2. An existing study reported that smokers have a higher rate of survival than non-smokers after cardiopulmonary resuscitation (CPR), which is referred to as smoker’s paradox [42]. Co-occurring medical problems, high SBP versus DBP among cocaine users. Scholars claimed that high SBP is the more pressing issue, especially concerning kidney patients [25-27]. The finding of this study shows that the importance of SBP vs. DBP are contingent upon the medical problems as shown in Figure 4. Cough is the worst health problem in this finding: if diabetic kidney patients use both cocaine and tobacco, they are likely to have both high SBP and high DBP, although SBP is slightly higher (41.23% vs. 38.93%). High SBP is highly associated with high respiratory, high prothrombin, high creatinine, and cardiopulmonary. The finding of this study shows that high SBP is associated with high BNP and low WBC, while high DBP is correlated with high respiratory, overweight, and shortness of breath. Especially, if overweight diabetic kidney patients use cocaine and tobacco, they are likely to have high DBP (25.00%) over SBP (13.37%). The same findings are observed patients with shortness of breath (19.47 vs. 10.30) and high respiratory (21.52 vs. 13.50). Although this finding shows that SBP and DBP have different degrees of association with various medical problems among African American diabetic kidney patients, DBP has higher impacts than SBP on various consequential comorbidities from cocaine and tobacco use. Lastly, different impacts of SBP and DBP are not clearly observed for high prothrombin, high creatinine, and cardiopulmonary.

**Discussion**

Using an apriori machine learning technique, this study investigated what kinds of co-occurring medical problems African American diabetic kidney patients will have if they only use cocaine, this study further drilled down to investigate the effect of combined tobacco and cocaine use and finally, how SBP and DBP are associated with these discovered medical problems from the use of tobacco and cocaine. The contributions of this study are as follows: Firstly, this study teased out cocaine users from tobacco users and ran analysis separately to observe independent effects of cocaine use on medical comorbidities. This finding showed that cocaine users have high SBP, cough, high prothrombin problems, among others. Secondly, in order to investigate the effects of using both tobacco and cocaine, the next analysis combined these two groups. The finding reveals that high respiratory and high BNP problems become top medical problems followed by high SBP. Third, this study investigated how SBP and DBP are differently afflicted with the medical problems derived from cocaine and tobacco use. While high SBP is associated with high BNP and low WBC, this study clearly shows that DBP has more impact on a variety of medical problems.

More specifically, being overweight, having shortness of breath and high respiratory problems are much more likely to be associated with high DBP than SBP. This finding is somewhat contradictory with existing studies, but we speculated such contradictions that our study categorized various medical problems and investigated each medical problem with SBP and DBP while existing studies investigated a single correlation such as: a correlation between cocaine and BP, or tobacco and BP or ESKD and BP. Lastly, unlike a cohort study or traditional research method that relies on existing studies, apriori does not assume prior knowledge, but searches all data and discovers most frequently co-occurring medical problems and discover new knowledge. Using this method, this study discovered new medical issues such as low RBC, WBC, cough, and high prothrombin which are not well recorded in existing studies. This study, however, has some limitations. First, the findings are based on the CHFDW, although it is arguably the largest electronic medical records (EMR) data, if diabetic patients are not in this database, they are not included in this analysis. Second, we used the 2012 diabetic patients, which was the most recent complete dataset in 2016. Although we don’t anticipate patients’ co-occurring medical problems abruptly year to year, we still caution the reader.
For future studies, it is recommended to include medications to observe whether gender or racial background affects the medications taken for the same medical issues. Furthermore, including medications into the analysis may be able to answer the low white blood cell count problem. To validate the finding, meta-study is recommended.

Acknowledgement

This project is supported by the College of Arts and Sciences, and the Center for Research Program Development and Enrichment at the University of Oklahoma. This work was conducted with data from the Cerner Corporation’s Health Facts data warehouse of electronic medical records. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Cerner Corporation. The authors greatly appreciate Codii Jones’ editing.

References


